

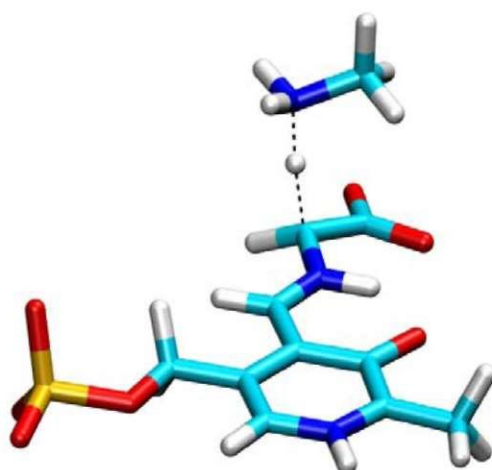


**Universitat de les Illes Balears**

*Departamento de Química*

# **Theoretical studies on pyridoxal 5'-phosphate-catalyzed reactions of biological relevance**

**Tesis doctoral**



**Rodrigo Casanovas Perera**

**2014**







**Universitat de les Illes Balears**

*Doctorado en Química Teórica y Modelización Computacional  
Departamento de Química*

# **Theoretical studies on pyridoxal 5'-phosphate-catalyzed reactions of biological relevance**

**Tesis doctoral**

**Autor: Rodrigo Casanovas Perera**  
**Director: Francisco Muñoz Izquierdo**  
**Director: Juan Frau Munar**

**Enero - 2014**



# **Theoretical studies on pyridoxal 5'-phosphate-catalyzed reactions of biological relevance**

A dissertation submitted in partial fulfilment of the requirements for the degree of  
Doctor of Theoretical Chemistry and Computational Modelling

Department of Chemistry  
University of the Balearic Islands

Presented by  
Rodrigo Casasnovas Perera  
2014

Thesis Advisor: Francisco Muñoz Izquierdo  
Thesis Advisor: Juan Frau Munar

Rodrigo Casasnovas Perera



El Dr. Francisco Muñoz Izquierdo, Catedrático de Química Física de la Universidad de las Islas Baleares y el Dr. Juan Frau Munar, Catedrático de Química Física de la Universidad de las Islas Baleares,

Certifican:

Que el presente trabajo de investigación, titulado "Theoretical studies on pyridoxal 5'-phosphate-catalyzed reactions of biological relevance", ha sido realizado bajo su dirección por D. Rodrigo Casasnovas Perera y constituye la memoria de su Tesis Doctoral

Palma de Mallorca, 2014

Firmado  
Dr. Francisco Muñoz Izquierdo  
Catedrático UIB

Firmado  
Dr. Juan Frau Munar  
Catedrático UIB

Directores de este trabajo





## Abstract

Pyridoxal phosphate (PLP) is a cofactor of more than a hundred enzymes that catalyze amino acid reactions like racemizations, transaminations and decarboxylations amongst others. After the formation of a Schiff base between PLP and the amino acid substrate, the mentioned reactions are favored by stabilization of a common carbanion species in the transition state. All the PLP-catalyzed reactions entail, at least, one step in which the C $\alpha$  carbon of the amino acid or the C4' carbon of the PLP is protonated or deprotonated. Furthermore, protonation of the carbanionic intermediate is the common crossroad to all possible reactions and it determines the final products. Since the experimental study of carbon acidities involves significant difficulties, several computational strategies were proposed in this work for the accurate prediction of pK<sub>a</sub> values. A methodology that provides pK<sub>a</sub>'s with uncertainties equivalent to experiment for carbon acids was obtained. Such methodology was applied to the pK<sub>a</sub> prediction of other functionalities and adapted to the calculation of stability constants of metal complexes with successful results. The activation energies of protonation and deprotonation reactions of C $\alpha$  by diverse enzymatic residues were calculated in order to obtain a general view of their kinetics in PLP-dependent enzymes. Finally, QM and QM/MM metadynamics simulations were carried out on the PLP-catalyzed decarboxylation of ornithine in Ornithine decarboxylase and in aqueous- and gas-phases. The obtained results provide a picture of how PLP-dependent enzymes control the specificity of the desired reaction by favoring specific protonation states of the PLP cofactor.



## Resumen

El piridoxal fosfato (PLP) es cofactor de más de un centenar de enzimas que catalizan reacciones sobre aminoácidos como racemizaciones, transaminaciones o descarboxilaciones entre otras. Después de la formación de una base de Schiff entre el PLP y el aminoácido sustrato, las reacciones mencionadas son catalizadas por la estabilización de un carbanión común en el estado de transición. Todas las reacciones catalizadas por PLP implican al menos una etapa de protonación o desprotonación del carbono C $\alpha$  del aminoácido o del carbono C4' del PLP. Es más, la protonación del intermedio carbaniónico es la encrucijada que une todas las posibles reacciones y determina los productos finales. Puesto que el estudio experimental de la acidez de átomos de carbono presenta muchas dificultades, en este trabajo se han propuesto diversas estrategias computacionales para la determinación precisa de  $pK_a$ , obteniendo una metodología que proporciona  $pK_a$ 's con incertidumbres equivalentes a las experimentales para los ácidos de carbono. Dicha metodología se aplicó con éxito a la predicción de  $pK_a$ 's de otros grupos funcionales y también se adaptó al cálculo de constantes de estabilidad de complejos metálicos. Las energías de activación de las reacciones de protonación y desprotonación de C $\alpha$  catalizadas por diversos residuos enzimáticos fueron calculadas con el fin de obtener una visión general de su cinética en enzimas PLP-dependientes. Por último, se realizaron simulaciones de metadinámica QM y QM/MM sobre la descarboxilación de ornitina catalizada por PLP en la enzima Ornitina descarboxilasa, en disolución y en fase gas. Los resultados obtenidos configuran una visión general de cómo las enzimas PLP-dependientes controlan la especificidad de la reacción deseada favoreciendo estados de protonación específicos del cofactor PLP.



## Agradecimientos

Desde que empecé los primeros trabajos de esta tesis, mucho ha pasado... y sin embargo ha sido sólo un momento. Pero no he estado solo en esta aventura. Muchas personas se han visto envueltas directa o indirectamente en la realización de esta tesis y merecen mi agradecimiento. Por lo que, para todos los que de algún modo os sentís partícipes: Muchas gracias.

En primer lugar y de forma especial debo agradecer a mis directores de tesis, el Dr. Francisco Muñoz Izquierdo y el Dr. Juan Frau Munar, su dedicación durante todo este tiempo, su implicación en el trabajo, su paciencia y sobre todo la oportunidad de realizar este trabajo con la máxima libertad posible.

También debo un gran agradecimiento al Prof. Michele Parrinello por acogerme de la mejor manera posible en su grupo de investigación cuando apenas sabía lo que era una simulación de dinámica molecular. Por su amabilidad, su buen humor y sus ¡Don Rodrigooo! También por su gran paciencia conmigo y sobre todo por la oportunidad de descubrir un mundo nuevo y fascinante. Grazie mille!

Además debo agradecer a muchos miembros del grupo de investigación. En primer lugar, debo agradecer a la Dra. Josefa Donoso su amabilidad conmigo, su interés e implicación en esta tesis. Recuerdo el día que gracias a ella un día recogí un libro de la biblioteca y la partícula en una caja me llevó al efecto túnel, al átomo de hidrógeno. Estoy convencido de que esta tesis empezó gracias a esa partícula en una caja.

A los Dres. Joaquín Ortega y Miquel Adrover por su colaboración y su participación en los trabajos aquí presentes. En particular, al Dr. Miquel Adrover porque tras apenas unos hablar unos minutos de unos resultados propuso y realizó él mismo unos experimentos que han resultado de gran importancia para esta tesis. También al Dr. Joaquín Ortega, mi compañero de despacho todo este tiempo, por las colaboraciones que han resultado tan importantes en esta tesis, por sus bromas y porque aguantarme en el mismo cuarto durante tantos años no habrá resultado fácil en muchas ocasiones.

Al Dr. Antoni Salva, por tutelarme en mis primeras aventuras con las matrices zeta, los cálculos semiempíricos en Mopac, Bader y la piridoxamina.

Al Dr. Bartolomé Vilanova por su amabilidad conmigo durante todo este tiempo y su atención e interés cuando le he planteado dudas sobre procedimientos experimentales que probablemente debería ya saber de la carrera.

A los becarios del grupo de Química Física que han hecho o están haciendo sus Tesis de Máster o de Doctorado durante este tiempo y con quienes he compartido muchas horas divertidas en la universidad, especialmente en la hora más importante del día: la comida. Gracias David, Carlos, Cati. Dra. Catalina Caldés, Carlos Maya, Laura, Marta, Jazmín, Christian, Jessica y muchos otros que en algún momento habéis compartido conmigo lo que yo llamo la buena y la mala vida de becario.

También debo agradecer a todos los compañeros del grupo del Prof. Michele Parrinello en Lugano. Muy especialmente a Daniela Wirz, quien me facilitó las dos estancias en Suiza hasta el infinito, por la forma de realizar su trabajo siempre de buen humor, con generosidad y amabilidad, y ante todo por hacernos sentir en casa.

Debo expresar un especial agradecimiento al Dr. Sebastiano Caravati por introducirme en CP2K y porque tras muchos intentos, conseguí compilarlo sin errores y acabar las metadinámicas QM/MM.

Además, hay muchas personas del grupo de Suiza a quienes debo agradecer: Dr. Gareth Tribello, Dr. Rustam Khaliullin, Dr. Michele Cerioti, Dr. Ali Hassanali, Dr. Vittorio Limongelli, Dr. Giacomo Miceli, Dr. Meher Prakash, Dr. Alessandro Barducci, Dr. Ivan Rivalta y muchos otros que fueron realmente buenos compañeros de los que aprendí mucho, compartí buenas vivencias y tengo los mejores recuerdos personales y científicos.

Debo a gradecer mucho a mis amigos de siempre y los no tan de siempre simplemente por eso, por ser mis amigos. Quiero hacer especial mención a Juan "Barce" por ser entre todos quién más ha sufrido las consecuencias de mi falta de tiempo libre y por aparecer inesperadamente para obligarme a volver al mundo real, a dar un paseo y hablar de las cosas que realmente son importantes ¡y de las que no son en absoluto importantes también, por supuesto!. También a Alex, Javi, David y Cris por muchos motivos... ¡qué bien lo pasamos cuando me vinisteis a visitar a suiza! Gracias a los compañeros de los cursos de doctorado, Jose, Ana, Elisa, Juan, Manu... todos "Masters", algunos ya Doctores y todos amigos.

Por supuesto, a mi familia, mis padres Eduardo y Cristina y mi hermana Susana: muchas gracias por todo durante todo este tiempo. Como ya dije una vez, ellos tienen gran responsabilidad en que yo esté aquí, en esta situación, en este momento y todo sin decirme nunca que me pusiera a estudiar. También a Charo, Ramón y Adrián, quienes siempre han sido generosos y excelentes conmigo y me hacen sentir parte de su familia.

Finalmente, quiero agradecer todo a Marta: por ser mi amiga, por ser mi amor, por su amor, por su bondad y generosidad, porque su mente me mantiene joven, por compartir su vida conmigo, por compartir mi vida con la Química y por muchas pequeñas y grandes razones que necesitan una vida para contarse.

A todos los nombrados y también a los innombrados, que sois igual de felices por no aparecer como los nombrados por hacerlo, gracias de nuevo a todos.







## Table of Contents

<b>1. Introduction.....</b>	<b>1</b>
1.1. Vitamin B6.....	3
1.2. Pyridoxal 5'-phosphate as a cofactor.....	3
1.3. Carbon acidity in the enzymatic reactions catalyzed by PLP.....	6
1.4. Acidity of PLP Schiffbases in aqueous solution.....	11
1.5. Reactions catalyzed by PLP in the presence of metal ions.....	16
1.6. Computational studies on PLP-catalyzed reactions.....	18
1.7. Vitamin B6 and the inhibition of glycation reactions.....	19
<b>2. Methodology.....</b>	<b>23</b>
2.1. The Schrodinger Equation.....	25
2.2. The Born-Oppenheimer Approximation.....	25
2.3. Pauli Exclusion Principle and Slater determinants.....	26
2.4. The Hartree-Fock Approximation.....	28
2.5. The Roothan-Hall equations.....	30
2.6. Form of the exact wave function and Electron Correlation.....	32
2.7. Foundations of Density Functional Theory, the Hohenberg-Kohn Theorems.....	34
2.8. The Kohn-Sham method.....	35
2.9. Exchange-correlation functionals.....	38
2.10. Continuum solvent models.....	39
2.11. Computational determination of $pK_a$ values.....	41
2.12. The CBS-QB3 method.....	44
2.13. Molecular Dynamics Simulations.....	45
2.14. Metadynamics simulations: Free Energy calculations and study of rare events.....	46
<b>3. Objectives.....</b>	<b>49</b>

<b>4. Results.....</b>	<b>55</b>
4.1. Evaluation of computational strategies for the calculation of $pK_a$ values and logp values of metal complexes.....	59
4.1.1. Simplification of the CBS-QB3 method for predicting gas-phase deprotonation free energies.....	61
4.1.2. Absolute and relative $pK_a$ calculations of mono and diprotic pyridines by quantum methods.....	71
4.1.3. Avoiding gas-phase calculations in theoretical $pK_a$ predictions.....	81
4.1.4. Theoretical calculations of stability constants and $pK_a$ values of metal complexes in solution: application to pyridoxamine-copper(II) complexes and their biological implications in AGE inhibition.....	97
4.2. Studies on the carbon acidities of PLP and PMP Schiffbases.....	111
4.2.1. Theoretical study on the distribution of atomic charges in the Schiffbases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms.....	113
4.2.2. C-H Activation in Pyridoxal-5'-phosphate Schiff Bases: The Role of the Imine Nitrogen. A Combined Experimental and Computational Study.....	123
4.2.3. C-H Activation in Pyridoxal-5'-phosphate and Pyridoxamine-5'-phosphate Schiff Bases: Effect of Metal Chelation. A Computational Study.....	137
4.3. Studies on the catalyzed Ca-C bond breaking and formation of PLP Schiffbases. . . . .	149
4.3.1. Non-enzymatic Pyridoxal 5'-Phosphate-catalyzed aldol condensation between amino acids and sugars. An inhibition mechanism of Advanced Glycation End-Products (AGEs) formation.....	151
4.3.2. Extraordinaire decarboxylation rates catalyzed by modestly efficient enzymes. A QM/MM metadynamics study on the enzymatic and nonenzymatic pyridoxal 5'-phosphate-catalyzed decarboxylation of amino acids.....	187
<b>5. Discussion.....</b>	<b>227</b>
5.1. Theoretical $pK_a$ calculations.....	229
5.2. PLP-catalyzed reactions.....	237
<b>6. Conclusions.....</b>	<b>249</b>
<b>7. Bibliography</b>	<b>253</b>



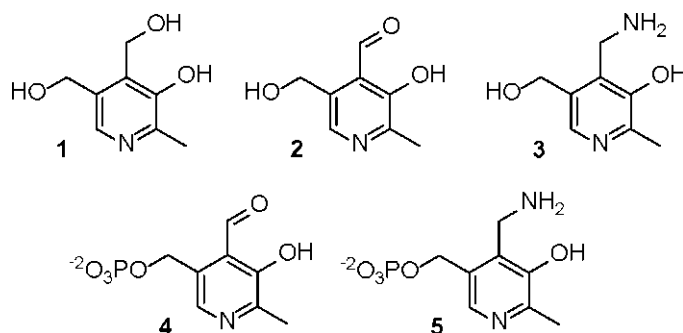


# **1. Introduction**



### 1.1. Vitamin B6

Vitamin B6 exists in three different molecular species, also known as vitamers, namely the alcoholic form (pyridoxine, PN) **1**, the aldehydic form (pyridoxal, PL) **2**, and the amino form (pyridoxamine, PM) **3** (Scheme 1). These species can be interconverted under physiological conditions, but it is worth to note that the biologically active forms are the pyridoxal 5'-phosphate (PLP) **4** and pyridoxamine 5'-phosphate **5**, which correspond to the phosphorylated derivatives at position 5' of pyridoxal and pyridoxamine (Scheme 1).



Scheme 1. Molecular forms of Vitamin B6.

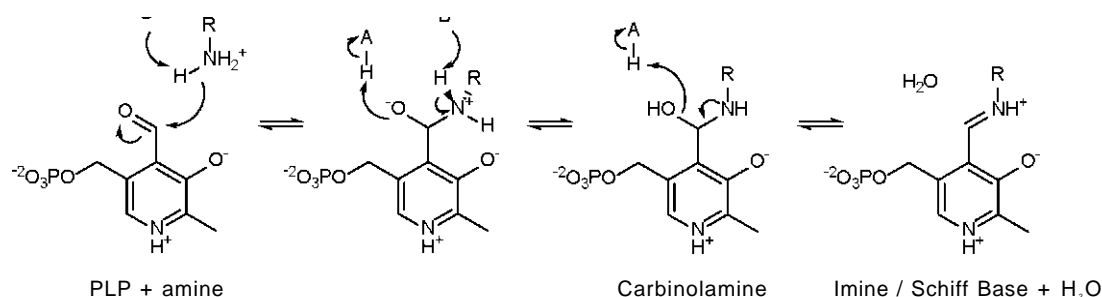
### 1.2. Pyridoxal 5'-phosphate as a cofactor

The spontaneous reactions involving the Cα bonds of amino acids in aqueous solution are amongst the slowest biological processes, some of which exhibit half-lives of as much as 1.1 billion years (Radzicka1996, Snider2000, Wolfenden2001). This may question the possibility of life formation when considering the age of the Earth (~4.5 billion years). However, taking into account the number of residues of each protein and the average number of proteins in the cell, such chemical lifetimes are necessary to prevent spontaneous degradation of proteins under physiological conditions (Wolfenden2001).

On the other hand, it is also required that all biological reactions proceed coordinately for the correct operation of the cell. Therefore, all reactions should take place in similar timescales, typically ps to ms, which shows the importance of catalysis for life. Many biological catalysts, also known as enzymes, are proteins which provide a most favorable environment for a specific reaction to take place between specific reactants. Some enzymes require the presence of a metal ion or an organic molecule, known as cofactor, to facilitate the catalysis.

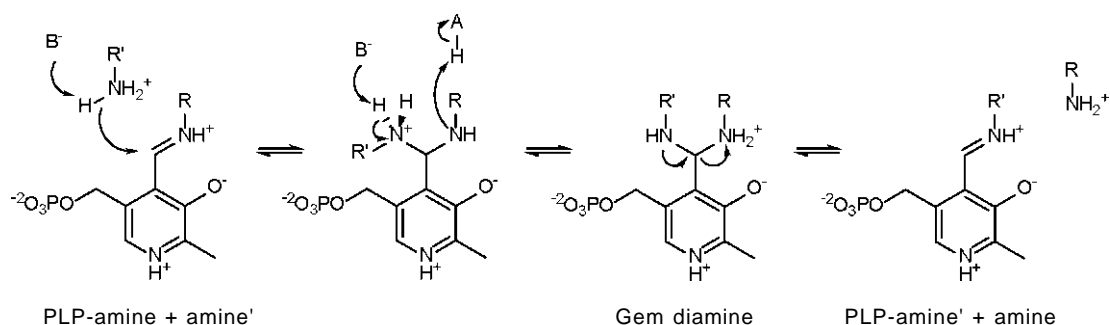
In the case of nitrogen metabolism, especially for amino acid reactions, the aldehydic form of Vitamin B6 or pyridoxal 5'-phosphate, PLP, is an essential enzyme cofactor. More than a hundred enzymes use PLP to catalyze transaminations, racemizations, α-decarboxylations, α-P- and γ-replacements and retro aldol cleavages of amino acids amongst other reactions (Evangelopoulos1984; Christen1985). In these

enzymes, the PLP reacts with the  $\alpha$ -amino group of a conserved lysine residue to form an imine adduct or Schiff base (Scheme 2). This PLP-Lys imine adduct, or Schiff base, is also commonly known as internal aldimine.



Scheme 2. Reaction mechanism of Schiff base formation between PLP and amines.

The first step in the PLP-catalyzed reactions of amino acids in enzymes is a transimination reaction, which consists in the substitution of the lysine linkage of the internal aldimine to form a new Schiff base with the  $\alpha$ -amino group of an amino acid (Scheme 3). This species is also known as external aldimine since it is formed with the incoming amino acid substrate in contraposition with the internal aldimine.



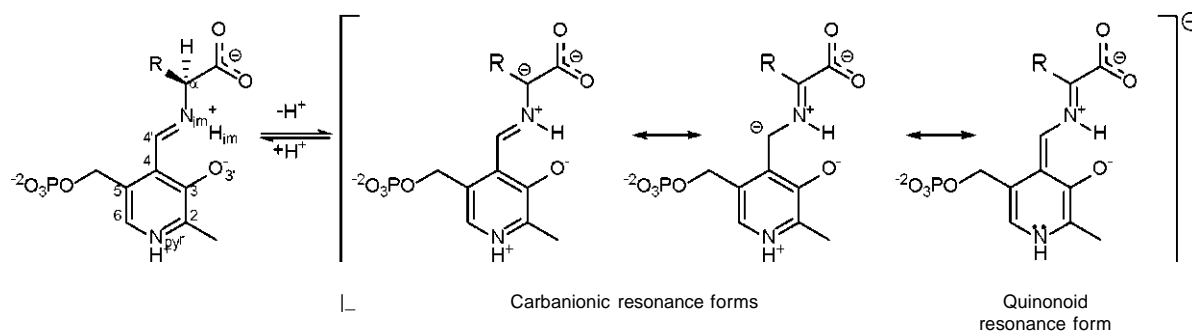
Scheme 3. Reaction mechanism of transimination of a PLP Schiff base and two amines.

It is worth to mention that the 5'-phosphate group does not participate in the catalytic process in PLP-dependent enzymes (Evangelopoulos1984, Christen1985). This group forms hydrogen bonds and/or salt bridges with polar and cationic groups of enzymatic residues or with the amide N-H hydrogens of the protein backbone. These interactions with the phosphate group also contribute to maintain the PLP cofactor in the correct position and orientation in the active site, in addition to the imine formed with the conserved lysine residue. However, the PLP Glycogen phosphorylase enzymes are an exception because the phosphate group of PLP acts as an acid-base catalyst (Evangelopoulos 1984, Livanova2006).

Once the PLP-amino acid external aldimine is formed, the next step is the heterolytic cleavage of one of the bonds of the alpha carbon of the amino acid, Ca, which generates a negative charge on such atom. The formation of a carbanion at the Ca position is favoured by the stabilization of the negative charge in the transition state



across the  $\pi$  system of the PLP-amino acid aldimine (Eliot2004, Jansonius1998, Toney2005) (Scheme 4).



Scheme 4. Schiff base formed between PLP and an amino acid and resonant forms resulting from proton abstraction at C $\alpha$ .

Dunathan (Dunathan1966) proposed, in a very elegant hypothesis, that the C $\alpha$ -bond to be cleaved should be oriented perpendicular to the pyridine ring of the Schiff base. This arrangement maximizes the overlap between the  $p$  orbital of the nascent negative charge at C $\alpha$  and the  $p$  orbitals of the  $\pi$  system in the transition state. As a result, the delocalization of such negative charge across the  $\pi$  system is also maximized, the energy barrier is lowered and the formation of the carbanion accelerated. It is important to note that this feature also provides a simple and efficient mechanism of controlling the reaction specificity. The hypothesis proposed by Dunathan is widely supported by crystallographic studies. The X-ray structures of PLP-dependent enzymes show that each active site favours a specific conformation of the C $\alpha$  substituents in which, with no exception, the bond that remains perpendicular to the plane of the external aldimine is cleaved (Eliot2004, Toney2011, Fogle2011). Studies of Toney and co-workers (Griswold2012, Spies2007) showed that hyperconjugation of the C $\alpha$ -H bond with the  $\pi$  system also reduces the activation barrier of proton abstraction due to a decrease of  $\sim 20\%$  in the bond order in the ground state.

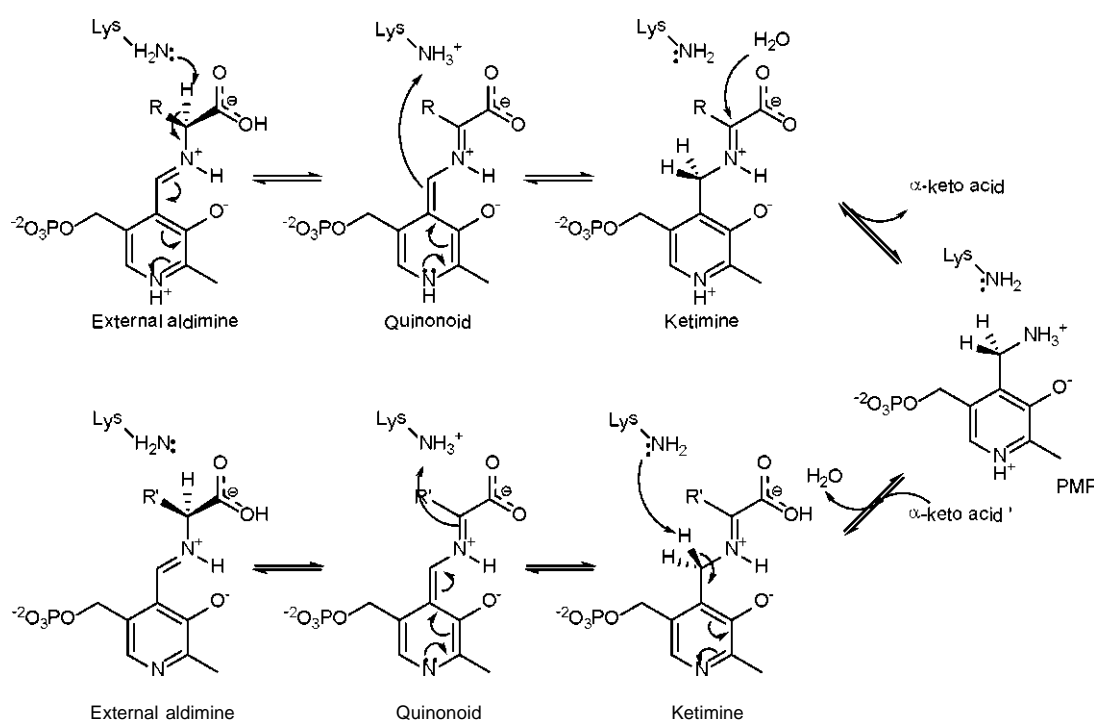
Apart from stereoelectronic effects, the protonation state of the heteroatoms of the external aldimine contributes to the stabilization of the negative charge in the transition state and in the formed carbanion intermediate (Evangelopoulos1984, Christen1985, Eliot2004, Toney2011). In many PLP-dependent enzymes, the pyridine nitrogen interacts with the acidic group of an aspartate or glutamate residues. The difference between in the acidities of the carboxylic group and pyridine nitrogen guarantees the protonation of the last group. The protonated pyridine nitrogen causes the complete delocalization of the negative charge formed at C $\alpha$  by the so-called "electron sink" effect (Scheme 4). As a consequence of this particular electronic distribution, this carbanionic species is also known as quinonoid intermediate (Scheme 4), and exhibits a characteristic UV-Vis spectrum, which is employed to detect its formation and monitor the course of PLP reactions (Evangelopoulos1984, Christen1985, Eliot2004)

### 1.3. Carbon acidity in the enzymatic reactions catalyzed by PLP

Some of the most common reactions catalyzed by PLP as a cofactor are introduced in this section to emphasize the importance of proton transfer reactions involving carbon atoms in PLP-dependent enzymes. Pyridoxal 5'-phosphate catalyzes a broad diversity of amino acid reactions, all of which include at least one proton transfer involving the C $\alpha$  or C4' atoms.

Transamination is one of the most studied enzymatic reactions catalyzed by PLP (Christen1985, Eliot2004). The accepted reaction mechanism is shown in Scheme 5. After the formation of the external aldimine, the first half reaction is initiated by proton abstraction at the C $\alpha$  carbon, generating the quinonoid intermediate. Then, the reaction evolves via protonation at C4' to produce a new class of Schiff base named ketimine, which is hydrolyzed by a water molecule yielding an  $\alpha$ -keto acid and the cofactor in the amine form, PMP. The second half of the reaction proceeds exactly oppositely. That is, a new ketimine is formed between PMP and a different  $\alpha$ -keto acid and the C4' carbon is deprotonated. Eventually, reprotonation at C $\alpha$  results in the aldimine form, which is hydrolyzed to regenerate the PLP cofactor and produce a new  $\alpha$ -amino acid (Scheme 5).

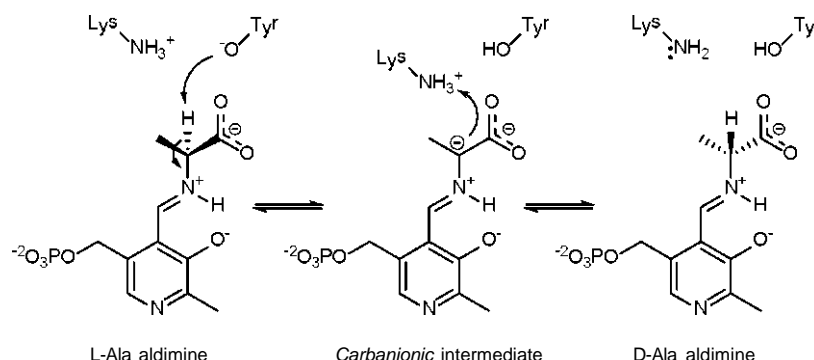
Note that such apparently complex mechanism is required to break the C $\alpha$ -N bond, which cannot be cleaved in a single step by the same mechanism as the rest of the C $\alpha$  bonds since the alpha nitrogen constitutes the imine linkage to the cofactor.



Scheme 5. Reaction mechanism of PLP catalyzed transamination.

Racemization reactions are similar to transaminations as both proceed via acid-base reactions involving the carbon atoms. The proposed reaction mechanism for Alanine racemase (AlaR) from *Bacillus stearothermophilus* is shown in Scheme 6

(Sun1999). Once the external aldimine is formed with L-alanine, the phenoxide anion of a tyrosine residue abstracts the  $\alpha$ -proton from the si-face yielding the carbanion intermediate. In the next step, the protonated  $\epsilon$ -amine group of a lysine residue reprotonates the  $\alpha$ -carbon on the re-face yielding the D-alanine aldimine (Scheme 6).

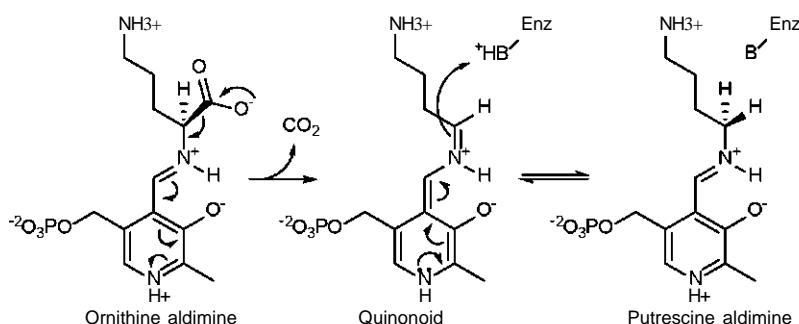


Scheme 6. Racemization mechanism of PLP-dependent Alanine racemase.

According to X-ray structures (Shaw1997, LeMagueres2005), the active site of AlaR exhibits a singularity amongst PLP-dependent enzymes because the pyridine nitrogen atom interacts with the positive guanidinium group of an arginine residue. Therefore, the pyridine nitrogen remains unprotonated, which avoids the electron sink effect and hinders the formation of the quinonoid intermediate in this reaction. In fact, spectroscopic analyses of alanine racemization by AlaR proved the absence of this resonant form during the course of the reaction (Spies2004). The most likely explanation for the prevention of the delocalization of the negative charge from the  $\alpha$ -carbon to the pyridine nitrogen is that by destabilizing the intermediate, its lifetime is reduced, which drastically reduces the possibility of side-reactions such as protonation at the C4' atom (Spies2004).

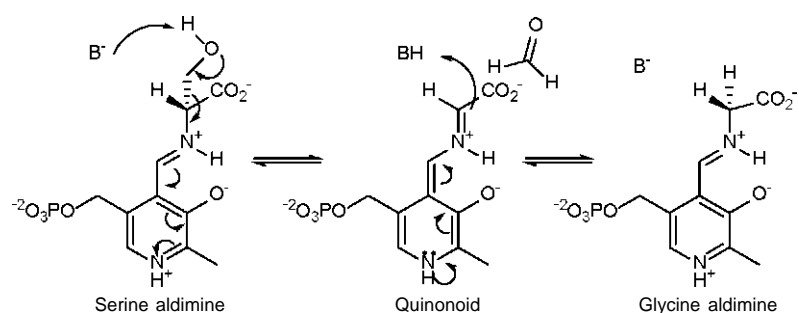
The reaction mechanism of decarboxylation in Ornithine decarboxylase (Jackson2000, Jackson2003) is depicted in Scheme 7. As in other PLP-dependent enzymes, the first steps of the reaction consist in the formation of the external aldimine. During the Schiff base formation reaction, the carboxylate group of the ornithine substrate is isolated from the water solvent in a hydrophobic pocket, which promotes the decarboxylation step (Jackson2003). Diaminopimelate decarboxylase catalyzes the decarboxylation of D,L-diaminopimelate in a completely equivalent mechanism to that shown for Ornithine decarboxylase in Scheme 7. However, it is very interesting to note that the carboxylate group in the PLP-diaminopimelate aldimine is not buried in a hydrophobic environment inside the active site but, contrarily, it is oriented directly towards the solvent. Furthermore, what is really unpredictable is that both enzymes exhibit similar reaction turnover numbers ( $k_{cat}$ ) as well as similar catalytic efficiency ( $k_{cat}/K_M$ ) (Fogle2011). The kinetic data of these enzymes, together with the conformation of the carboxylate groups in their respective active sites indicates that the catalytic origins cannot be exclusively attributed to destabilization of the negative charge in the ground state by an hydrophobic environment in the active site (Fogle2011).

Nevertheless, in both decarboxylase enzymes, the arrangement of the carboxylate group in the active site labilizes the Ca-COO<sup>-</sup> bond by being oriented perpendicularly to the  $\pi$  system (Jackson2000, Jackson2003, Fogle2011). Once the CO<sub>2</sub> is eliminated, an acidic residue transfers a proton to the Ca atom to yield the aldimine product (Scheme 7).



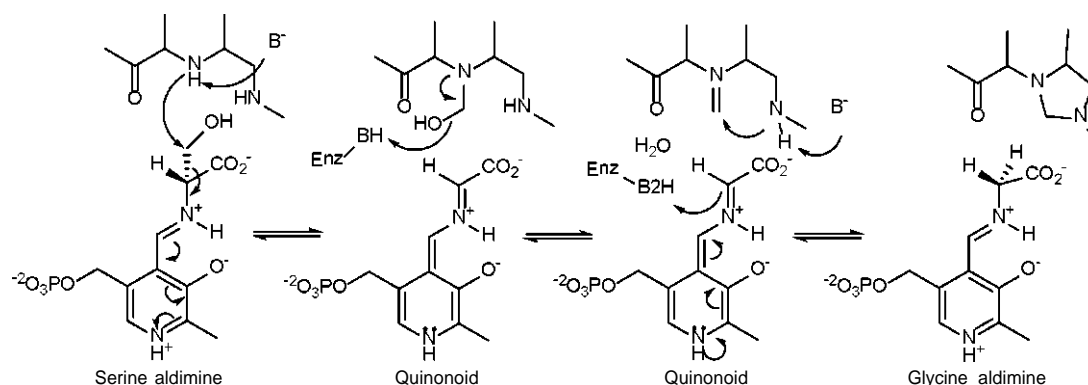
Scheme 7. Reaction mechanism of PLP-dependent Ornithine decarboxylase.

The last sort of PLP-catalyzed reactions at the  $\alpha$  carbon of amino acids are a-eliminations and replacements (Eliot2004, Toney2011). Serine hydroxymethyl transferase catalyzes a reversible retro-aldol reaction between serine and glycine. To date there is still controversy about the reaction mechanism of the catalyzed reaction (Szebenyi2004, Schirch2005). A simple retro-aldol mechanism is shown in Scheme 8, the retro aldol cleavage is initiated by deprotonation of the  $\beta$ -hydroxyl group of serine in the aldimine adduct, which promotes the elimination of C $\beta$  in the form of formaldehyde and the formation of the quinonoid intermediate. In the next step, the quinonoid intermediate is protonated at the C $\alpha$  atom, yielding a glycine aldimine which is eventually hydrolyzed.



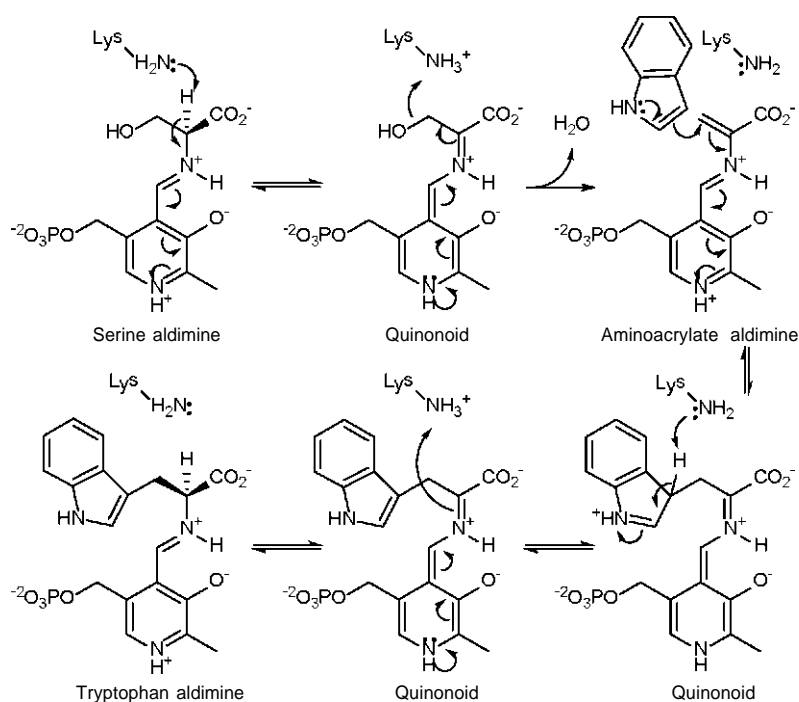
Scheme 8. Proposed retro-aldol reaction of serine aldimine.

Crystallographic structures support an alternative reaction mechanism that involves the participation of tetrahydrofolate as a carrier between serine and glycine (Szebenyi2004, Schirch2005) (Scheme 9). Instead of a retro-aldol cleavage, the reaction starts by a nucleophilic substitution on C<sub>P</sub> by the attack of the tetrahydrofolate co-factor, which causes the cleavage of the Ca-CP bond and generates a quinonoid intermediate. Subsequently, the quinonoid intermediate is protonated at C $\alpha$  to yield a glycine aldimine (Scheme 9).



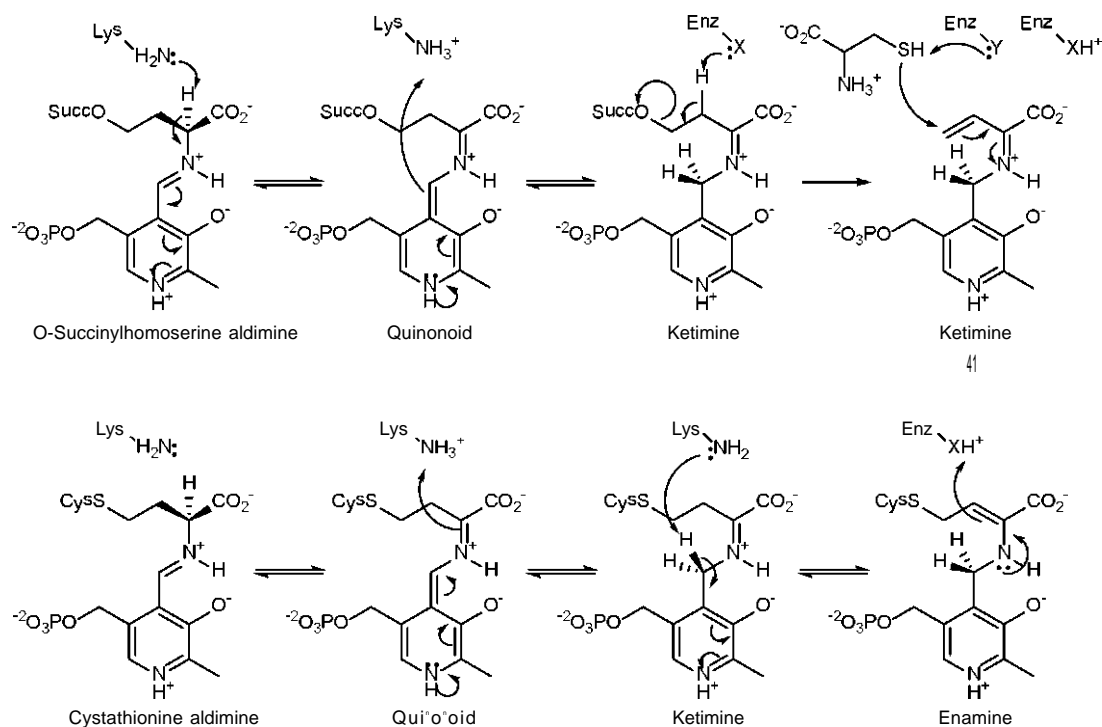
Scheme 9. Central steps in the mechanism of  $\alpha$ -elimination and replacement catalyzed by Serine hydroxymethyltransferase. The BH and B2H labels stand for enzyme residues that act as acid catalysts in the course of the reaction, which are still unidentified.

Additionally, PLP-dependent enzymes also catalyze reactions at P- and  $\gamma$ -carbons of amino acids. Tryptophan synthase catalyzes serine conversion to tryptophan via a P-elimination and replacement reaction (Miles2001). In the first step, the  $\alpha$ -carbon is deprotonated by a lysine residue to generate a quinonoid intermediate (Scheme 10). Subsequently, the protonation of the P-hydroxyl group of serine favours its elimination, yielding an aminoacrylate aldimine. The Michael-type addition of an indol group to the CP carbon generates a new quinonoid intermediate which is reprotonated at C $\alpha$  to yield the tryptophan aldimine (Scheme 10).



Scheme 10. Reaction mechanism of  $\beta$ -elimination and replacement catalyzed by Tryptophan synthase.

The reaction mechanism of  $\gamma$ -elimination and replacement in Cystathionine  $\gamma$ -syntase (Eliot2004, Brzovic1990) is shown in Scheme 11. After the formation of the external aldimine, a lysine residue deprotonates the C $\alpha$  carbon. Then, as in the transamination reaction, a ketimine Schiff base results from protonation of the quinonoid intermediate at the C4' carbon. This species forms an  $\alpha$ ,  $\beta$ -unsaturated imine by proton abstraction at C $\beta$ , in which the Cy substituent is simultaneously eliminated. The unsaturated ketimine undergoes Michael addition at Cy to complete the replacement reaction. Subsequent protonation at C $\beta$  regenerates the ketimine, which later evolves to a quinonoid intermediate by deprotonation of the C4' carbon. Eventually, the aldimine product results from protonation at C $\alpha$  atom.



Scheme 11. Reaction mechanism of  $\gamma$ -elimination and replacement catalyzed by Cystathionine  $\gamma$ -syntase.

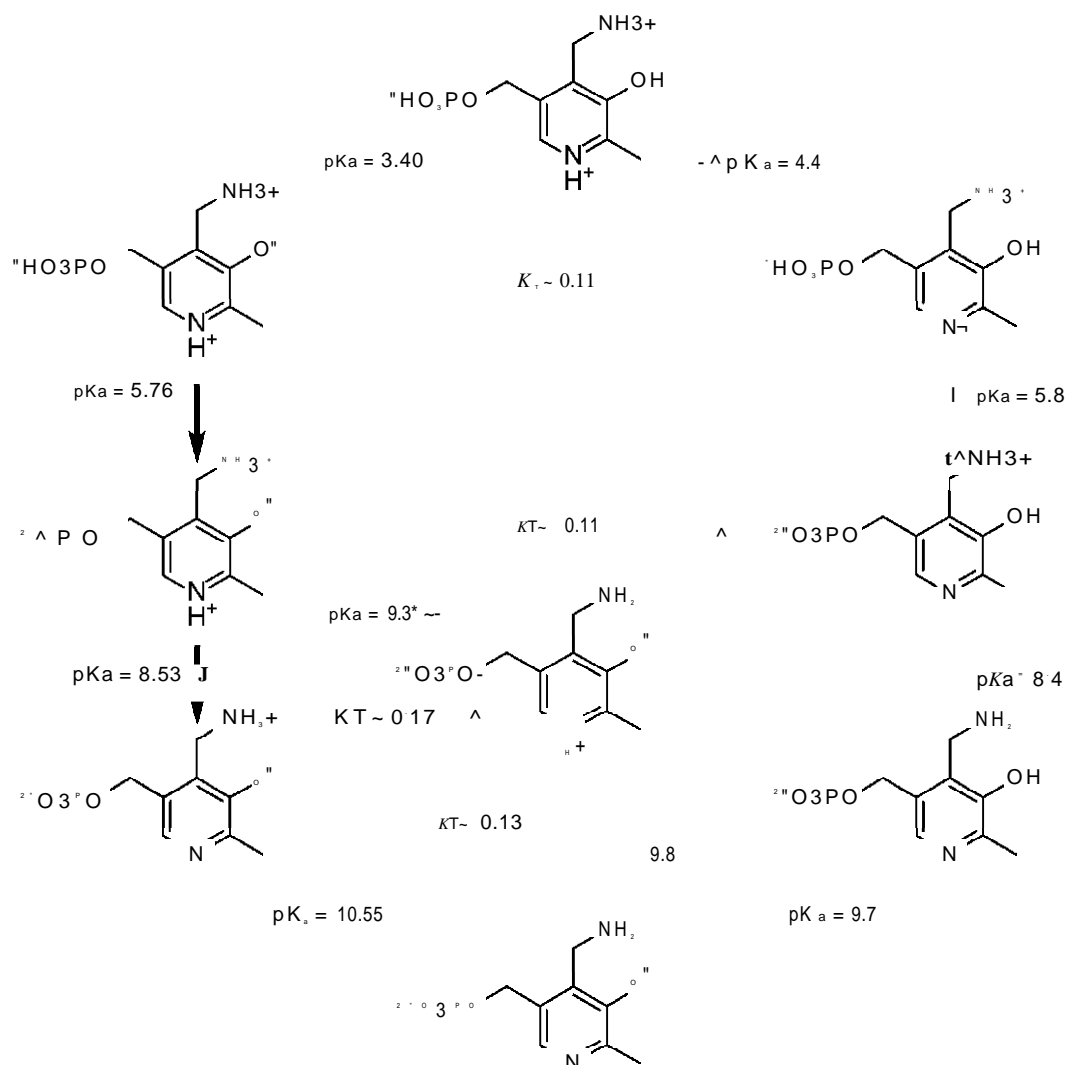
#### 1.4. Acidity of PLP Schiff bases in aqueous solution

As illustrated in the previous section, the protonation state of the external aldimine Schiff bases strongly affects the stabilization of the carbanion intermediates and their evolution towards protonation at the C $\alpha$  or C4' carbons. Therefore, knowledge of the acid-base chemistry of PLP and the Schiff bases formed with amino acids is essential to understand the catalysis in enzymes. Because of the fact that PLP-dependent enzymes exhibit active sites with numerous polar and charged residues, the reactivity of PLP Schiff bases in aqueous is significantly representative of the enzymatic chemistry.

The acid-base chemistry in aqueous solution of pyridoxal, pyridoxamine and their phosphorylated counterparts has been extensively studied (Evangelopoulos1984, Christen1985, Vilanova2004, Chan-Huot2010). Since PL, PM, PLP and PMP have several protonable groups, each vitamer exhibits several tautomeric equilibria in a wide pH range.

For example, the dissociation and tautomeric equilibria for PMP are depicted in Scheme 12. The first macroscopic  $pK_a$  is assigned to the first deprotonation of the phosphoric acid group since at pH  $\sim$ 2 the phosphate group bears a negative charge. Therefore, its first  $pK_a$  is estimated to be inferior to 2.5. Next, and considering the most abundant tautomers, the second acid dissociation corresponds to the phenol group with a  $pK_a$  value of 3.40. This group is surprisingly acid in comparison with the 3-hydroxypyridine phenol ( $pK_a=8.75$ ), which is due to the stabilization of the phenoxide anion by the protonated pyridine nitrogen and by hydrogen bond interaction with the protonated amino group. The third ionization of PMP is assigned to the second deprotonation of the phosphate group ( $pK_a=5.76$ ). The two groups that are deprotonated under basic pH conditions are the pyridine and amine groups, with  $pK_a$  values of 8.53 and 10.55 respectively (Scheme 12). It is worth noting that the tautomers which present ionized pyridinium and phenoxide groups are favoured in aqueous solution, while in non polar solvents the protonated phenol and deprotonated pyridine nitrogen tautomers are more stable (Chan-Huot2010).

Except for the deprotonation of the amine group, the exchange of this functionality for an aldehyde group at C4' has little effect on the acid-base chemistry of PLP with respect to PMP (Christen1985, Vilanova2004, Vazquez1989). The first  $pK_a$  value of the phosphate group is estimated to be less than 2.5 units. The phenol group shows a  $pK_a$  value of 3.28, which is very similar to that found for PMP even considering that phenoxide anion cannot be stabilized by hydrogen bonding interactions with the aldehyde group. The second deprotonation of the phosphate group presents a  $pK_a$  value of 6.1 units, and the last  $pK_a$  corresponds to the pyridine nitrogen with a value of 8.33. Similarly to PMP, the zwitterionic and neutral tautomers are respectively more abundant in aqueous solution and in non polar solvents. Despite all similarities with PMP, the aldehyde PLP form exhibits hydrates at low pH in aqueous solution resulting from water addition to the C4' carbon (Chang-Huot2010).



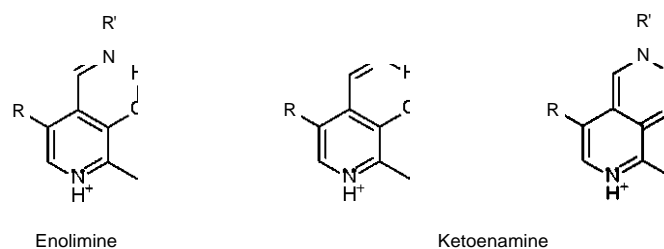
Scheme 12. Acid base and tautomeric equilibria for PMP.

Some dissociation constants vary several  $\text{pK}_a$  units when the Schiff bases are formed (Christen1985, Vazquez1989, Vazquez1990, Vazquez1991, Vazquez1992, Chang-Huot2010). The acid-base behaviour of the 5'-phosphate group remain almost invariant in relation to PMP or PLP after Schiff base formation (Scheme 13). In fact, the first dissociation corresponds to the phosphate group with an acidity constant lower than  $\text{pK}_a$  2, while  $\text{pK}_a$  of the second proton of the phosphate group is 5-6. On the other hand when Schiff bases are formed between PLP and amino acids, a new dissociation corresponding to the carboxylic acid is measured with  $\text{pK}_a \sim 2$ . The phenol and pyridine groups show  $\text{pK}_a$  values of 2.8 and 6.5 respectively, which shows an increase of acidity in relation to their PMP and PLP counterparts whereas the iminium group, with  $\text{pK}_a$  11-12 is more basic than its equivalent amine in PMP (Scheme 13).



The main tautomeric equilibrium in pyridoxal and pyridoxamine is the proton exchange between phenolic oxygen and pyridine nitrogen. However, the PLP Schiff bases present an additional sort of tautomerism in which the proton of the O3' phenol oxygen is exchanged with the imine nitrogen (Scheme 13, Scheme 14). This equilibrium is sensitive to the polarity of the environment and to hydrogen bond microsolvation (Chang-Huot2010). The predominant tautomer in aqueous solution, and other polar solvents that form hydrogen bonds with the phenol oxygen, show an unprotonated phenoxide anion and a protonated iminium cation groups (ketoenamine tautomers), while in non polar solvents the most abundant form has a protonated phenol

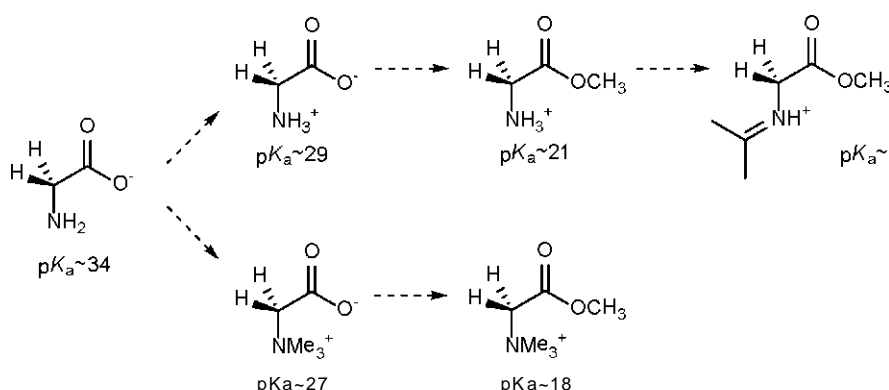
oxygen and an unprotonated imine nitrogen (enolimine tautomers) (Chang-Huot2010, Christen1985) (Scheme 14).



Scheme 14. Tautomeric forms of Schiff bases formed between PLP and amines.

A most interesting acid base behaviour of the PLP and PMP Schiff bases in relation to their catalysis, involves the C<sub>α</sub> and C<sub>4'</sub> atoms. The experimental determination of carbon acidities is complicated because of the weak acidities of C-H hydrogens, which typically exhibit pK<sub>a</sub> values larger than 20 in aqueous solution. However, the corresponding carbon acidities of several amino acids, peptides and other related compounds of biological relevance have been measured by NMR methods (Rios1997, Rios2000, Rios2001, Richard2002, Rios2002, Toth2007, Crueiras2008, Crueiras2009, Richard2009, Crueiras2011).

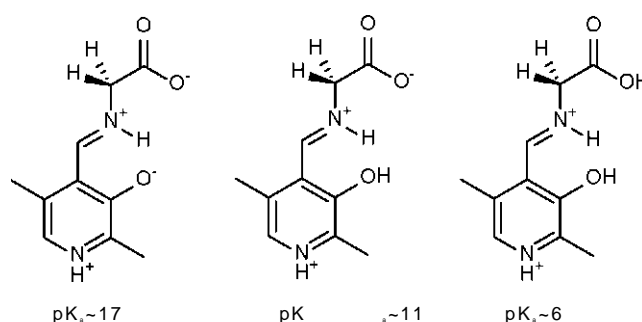
The C-H hydrogen of anionic glycine shows pK<sub>a</sub> 34 in aqueous solution, while protonation of the amino group to yield zwitterionic glycine increases the carbon acidity to pK<sub>a</sub> 29 because of the electrostatic stabilization of the carbanion by the ammonium cation (Rios2000, Rios2002) (Scheme 15). However, solvation of the protonated amine by water reduces the positive charge on the nitrogen atom and the electrostatic stabilization is not completely achieved. In fact, full methylation of the amine nitrogen increases the carbon acidity by a factor of 10 (pK<sub>a</sub>~27) with respect to the protonated amine (Rios2002) (Scheme 15)



Scheme 15. Substituent effects on the C<sub>α</sub> acidity of glycine in aqueous solution.

Once the proton is eliminated from C<sub>α</sub>, the resulting negative charge is delocalized via n-n delocalization to the carboxylate neighbour yielding an enolate anion. Therefore, protonation state of this group has also a significant effect on the carbon acidity of C<sub>α</sub>, which is exemplified by the reduction of 8-9 units in the pK<sub>a</sub> in

the methyl esters of glycine (Rios2002). The effects of Schiff base formation on the carbon acidity of Ca were reported by Richard and co-workers (Rios2001). Imine formation between glycine methyl ester and acetone increases the carbon acidity of Ca by a 10 factor (Scheme 15). A later study illustrated the power of 5'-deoxypyridoxal in enhancing the carbon acidity of glycine when forming Schiff bases (Toth2007). As shown in Scheme 16, the  $pK_a$  of Ca were reported only for three protonation states in aqueous solution. With respect to glycine zwitterion (i.e.  $pK_a \sim 2.9$ ), Schiff base formation with 5'-deoxypyridoxal reduces the  $pK_a$  of Ca by 12 units in the ketoenamine pyridinium form. However, further protonation of the Schiff base heteroatoms leads to  $pK_a$  values for the Ca atom which are typical of moderately strong acids.

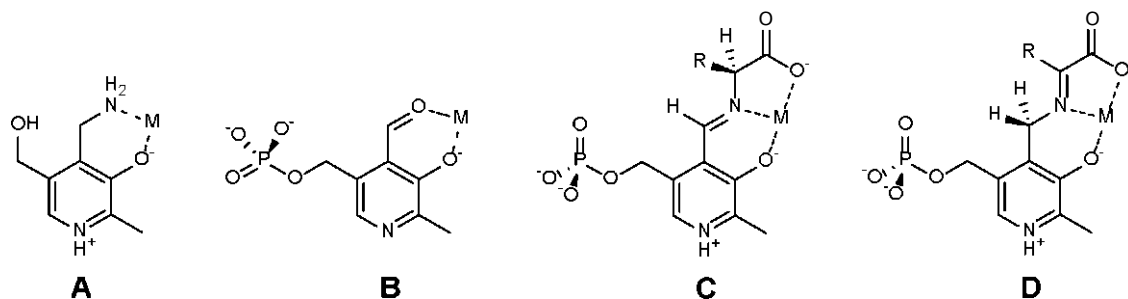


Scheme 16. Carbon acidities for different protonation states of 5'-deoxypyridoxal and glycine Schiff bases.

Nevertheless, much less research has been carried out to report the carbon acidities of pyridoxal Schiff bases in enzymatic media. Toney and co-workers obtained the reaction free energy profile for the racemization process in Alanine racemase of *Bacillus stearothermophilus* (Spies2004). As explained, the pyridine nitrogen of the Schiff base remains deprotonated in this enzyme, dramatically decreasing its electron-sink stabilizing properties. Accordingly, no quinonoid resonance form is detected spectroscopically in this enzyme (Sun1999, Spies2004). The free energy of deprotonation of Ca in the active site was estimated with a wide uncertainty to be between 4 and 12 kcal/mol. In this reaction, the proton is removed from Ca by the phenoxide anion of Tyr265', whose estimated  $pK_a$  is 7.2. From these values, the  $pK_a$  value of Ca in the active site is 10-16 (Sun1999, Spies2004).

### 1.5. Reactions catalyzed by PLP in the presence of metal ions

Pyridoxal, pyridoxamine and their phosphorylated derivatives form stable coordination complexes with a number of divalent and trivalent metals like Cu(II), Zn(II), Ni(II), Mn(II) and Al(III) amongst others (Leussing1986) (Scheme 17). Similarly, the Schiff bases of PLP and PMP chelate metal ions but acting as tridentate ligands (Leussing1986, Christen1985). In these complexes, the metal ion replaces the proton that is shared by the imine and phenol groups in the uncomplexed Schiff bases (Scheme 17).



Scheme 17. Pyridoxamine (A), Pyridoxal 5'-phosphate (B), Pyridoxal 5'-phosphate Schiffbase (C) and Pyridoxamine 5'-phosphate Schiffbase (D) metal complexes.

Despite the absence of metal complexes in PLP-dependent enzymes, the complexed Schiff bases are of high interest as model systems due to the similar reactivity with the Schiffbases in the enzyme active sites (Christen1985, Leussing1986, Martell1989). From a practical perspective, the use of metal ions is convenient in the study of PLP-catalyzed reactions for various reasons. Firstly, the complexation of metal ions stabilizes the Schiffbases and displaces their equilibrium of formation. As a result, the concentration of Schiff base is increased in solution facilitating their detection and monitoring during the catalysis on amino acids (Martell1989). Secondly, the stability of the Schiffbase complexes also increases in a broad range of pH, which allows the study of the catalysis for a richer variety of protonation states (Martell1989). Finally, the interaction of the metal ion modifies the spectroscopic properties of the Schiff base ligands, which facilitates or makes possible the monitoring of some reactions (Weng1983).

The transamination reaction kinetics of the PLP-ethylamine Schiff base with alanine or aspartate in the presence of Zn(II) was studied by Weng and Leussing (Weng1983). Their results show that either the phosphate or the phenol groups are intramolecular catalysts of the proton transfer between the amino groups in the gem diamine intermediate as the presence of buffer catalysts in solution does not modify the kinetics of the reaction. However, the reported rate constants do not clarify whether the Zn(II) ion catalyzes the transamination reaction because the experiments were deliberately design to minimize the presence of PLP-ethylamine-Zn(II) complexes. (Weng1983).

Concerning the Ca-H activation as a consequence of metal quelation, significant fraction of the PLP-alanine Schiffbase complexed with Al(III) becomes deprotonated in

solution to the point that the carbanionic intermediates are measurable by NMR methods (Martell1989). This result shows a significant activation of the Ca by Al(III), as well as the stabilization of the carbanionic intermediates with respect to protonation by water. In addition, equal concentrations of PLP-alanine and PMP-pyruvate Schiff bases were measured at equilibrium when the PMP-pyruvate-Al(III) complexes were used as starting reactants, which shows that chelation of Al(III) also activates the C4'-H hydrogens (Martell1984).

It is important to note that the degree of activation of the Schiff base ligand depends on the specific complexed metal ion. For example, the completely deprotonated Schiff bases of PMP and pyruvate complexed with Cu(II) undergo spontaneous transamination (Leussing1986). However, in the Zn(II) complexes, the reaction is only spontaneous on condition that the pyridine nitrogen becomes protonated (Leussing1986). Additionally, the monoprotonated PMP-pyruvate-Zn(II) complex is more reactive than the deprotonated free PMP-pyruvate Schiff base but less reactive than the monoprotonated free PMP-pyruvate Schiff base (Leussing1986). Therefore, apparently, complexation of Zn(II) does not catalyze transamination due to real ligand activation but due to an increase in the concentration of the Schiff base in solution.

Nevertheless, the experiments of Zabinski and Toney (Zabinski2001) show that the rates of the Ca deprotonation step in the complexes of Al(III) are somewhat slower, approximately 0.8-fold, than those in the free Schiff bases. Accordingly, the apparent reactivity enhancement in the metal complexes should only be attributed to an increase in the concentration of the Schiff bases rather than to real ligand activation.

As depicted in Scheme 17, the amino acid carboxyl group binds the metal ion together with the imine and phenol groups. As a result, the Ca-COO bond is fixed in the molecular plane of the imine and pyridine moieties. Therefore, decarboxylation reactions are prevented because in a hypothetical cleavage of the Ca-COO bond, the negative charge could not be stabilized by delocalization across the  $\pi$  system. Additionally, the negative charge of the carboxyl group is electrostatically stabilized on the metal ion by the coordinative interaction, which further prevents the loss of CO<sub>2</sub> (Martell1989, Zabinski2001). On the other hand, the rotation of the carboxyl group in the Schiff base complexes disposes the amino acid CP of the sidechain perpendicularly to the molecular plane and favours retro-aldol reactions in the CP hydroxylated amino acids such as serine.

### 1.6. Computational studies on PLP-catalyzed reactions.

The chemical reactivity PLP Schiff bases in aqueous solution and in enzymes has been studied from different computational approaches. The tautomerism between the phenol and pyridine groups in PLP and PLP-related species was studied in aqueous solution and non polar solvents (Kiruba2003). The combination of DFT/B3LYP and MP2 methods with discrete and continuum solvent approaches provided the correct tautomeric behaviour in polar and non polar solvents with respect to the experimental results. Accordingly, the neutral species resulted more stable in non polar solvents whereas the zwitterionic ones are predominant in aqueous solution. It is worth to note that the hybrid discrete-continuum solvation approach is required to obtain the correct tautomeric energies.

The mechanisms of Schiff base formation between pyridoxal and amines and transimination between methylamine and the PLP-methylamine Schiff base were computationally studied in aqueous solution with Density Functional Theory methods (Salva2001, Salva2002, Salva2003, Salva2004). Firstly, these works highlight the importance of the protonation state of PLP in aqueous solution to promote the Schiff base formation. Secondly, a mechanistic implication, with importance in enzymatic reactivity, is that reactive water molecules are required during the transimination, Schiff base formation and hydrolysis to catalyze the proton transfer reactions between the attacking nucleophile and leaving groups.

A QM/MM (ONIOM) study of the Schiff base formation in the active site of Ornithine decarboxylase showed that the reaction mechanism is analogous to the reaction in solution (Oliveira2011). In the active site, the thiol group of a cysteine residue, instead of a water molecule, acts as an acid/base catalyst transferring protons between the attacking amine nucleophile and the hydroxyl group of the carbinolamine intermediate (Oliveira2011). These results indicate that PLP-dependent enzymes also catalyze the Schiff base formation and hydrolysis in addition to transformations of the amino acid substrate. A cluster model of the active site of Ornithine decarboxylase was used for the study of the transimination reaction with Density Functional Theory calculations (Cerqueira2011). According to the obtained free energy profiles for different reaction pathways, the most favourable mechanism involves a proton exchange reaction between the incoming and leaving amino groups which is catalyzed by a water molecule. This study highlights the usefulness of investigating model reactions of PLP in aqueous solution.

The decarboxylation of different Schiff bases formed between amino acids and glyoxal or pyridoxal were studied in gas phase by using DFT/B3LYP and MP2 methods (Bach1997, Bach1999). An important conclusion of these studies is that the transition state is stabilized by the iminium cation adjacent to the developing charge at C $\alpha$ . Later, PM3 semiempirical studies were carried out on the decarboxylation reactions of the PLP-2-aminoisobutyrate in gas phase, in solution and in a model active site of Dialkylglycine decarboxylase (Toney2001). These computations also support that the stabilization of the transition state by the iminium cation is larger than that of the pyridine ring.

Lin and Gao (Lin2011, Lin2010) studied the decarboxylation of L-Dopa catalyzed by PLP in the active site of L-Dopa decarboxylase and in aqueous solution for the ketoamine and enolimine tautomers of the PLP-Dopa Schiff base. Their results suggest that in the active site, as well as in aqueous solution, the enolimine species is the most abundant tautomer (Lin2010). In addition, the enolimine tautomer reduces the free energy barrier of the deprotonation reaction in a larger extent than the ketoamine tautomer.

The racemization of alanine catalyzed by PLP in aqueous solution and in the active site of Alanine racemase was studied by QM/MM simulations. From umbrella sampling simulations and weighted histogram analysis techniques, a free energy barrier of 18.7 kcal/mol was obtained for the deprotonation of Ca by the phenoxide group of Tyr265' (Major2006, Major2006/2). The calculated reaction free energy in the simulations was 6.6 kcal/mol. From this result, the  $pK_a$  of the Ca carbon in the PLP-Ala Schiff base was estimated to be 12.2 (Major2006/2). In these studies, the carbon acidity enhancement in the non-protonated pyridine nitrogen aldimines is ascribed to stabilization of the carbanion intermediate by specific interactions between the imine moiety and water molecules in aqueous solution or specific groups in the enzymatic active site (Major2006, Major2006/2).

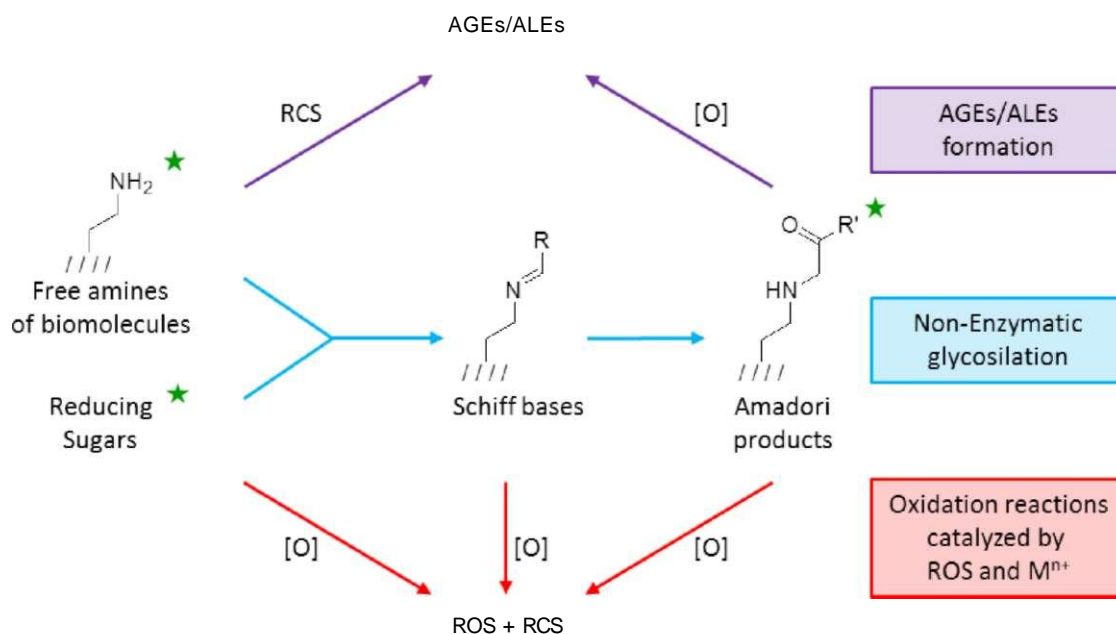
## 1.7. Vitamin B6 and the inhibition of glycation reactions

Some biomolecules, such as proteins and lipids amongst others, require the binding of glycans, which chemically are oligosaccharides or polysaccharides, to accomplish their function. The binding reactions, or glycosylations, are controlled and catalyzed by specific enzymes in the cellular environment to avoid the indiscriminate modification of biomolecules. However, abnormally elevated concentrations of reducing sugars or reactive carbonyl species (RCS) lead to the non enzymatic glycosylation, also known as glycation, of amino groups of proteins, aminophospholipids and nucleic acids (Rabbani2012, Li2008, Miyazawa2012).

The resulting product from the condensation between amino groups and sugars is a Schiff base adduct which may undergo an isomerization reaction yielding an Amadori compound or 1-imine-1-ketose species (Maillard1912) (Scheme 18). The inconveniences of uncontrolled glycation reactions arise from the high susceptibility of Schiff bases and Amadori compounds to oxidation reactions, which entail the degradation of the biomolecules to which are linked (Thornalley1984, Rabbani2012, Li2008, Miyazawa2012). Ultimately, Schiff bases and Amadori compounds degrade to advanced glycation and lipoxidation end-products, AGEs and ALEs respectively (Scheme 18).

Apart from the presence of sugars and reactive carbonyl compounds, some species such as hydroxyl, hydroxyperoxyl radicals (i.e OH<sup>-</sup> and HOO<sup>-</sup>), also known as reactive oxygen species (ROS), accelerate the degradation of glycated biomolecules due to their oxidizing potential. In addition, trace concentrations of free transition metals with high redox activity such as the pairs Fe(III)/Fe(II) and Cu(II)/Cu(I) catalyze the

formation of ROS in the Fenton reaction and, therefore, cause the degradation of glycosylated biomolecules as well (Kepp2012, Jomova2010) (Scheme 18).



Inhibition mechanisms of AGEs/ALEs formation by PLP and PMP:

- \* PLP/PMP Schiff base protection
- ROS scavenging
- \**k* Metal chelation
- \* RCS scavenging

Scheme 18. Formation routes of advanced glycation and lipoxidation end-products (AGEs/ALEs). The labels mark the compounds that react with PMP or PLP inhibiting the formation of AGEs and ALEs.

From the medical perspective, glycation of biomolecules and especially formation of AGEs and ALEs have been related to a number of medical conditions comprising ageing, diabetes, atherosclerosis, tissue degradation, inflammatory diseases (Brownlee1992, Kume1995, Bailey1998, Baynes1999, Grillo2008, Ramasamy2005) and neurodegenerative illnesses such as Alzheimer's and Parkinson's diseases (Hoyer2002, Miranda2010).

Vitamin B6 exhibits high inhibitory activity of AGE and ALE formation as it reacts and neutralizes the activity of the compounds responsible of the different routes of glycation and degradation of biomolecules. On one hand, the pyridoxamine form (PM) is highly reactive towards condensation with carbonyl groups of reducing sugars and RCS, yielding stable Schiff bases and reducing the initial steps of glycation (Adrover2005, Adrover2007, Adrover2009, Ortega-Castro2010, Voziyan2002, Voziyan2005). The aldehyde form of vitamin B6, PLP, also forms stable Schiff bases but with the free amino groups of proteins and lipids, which contributes to prevent their glycation by reducing sugars and RCS (Caldes2011) (Scheme 18).



Pyridoxamine also protects the carbonyl groups of Amadori compounds by forming Schiff bases once the glycation reactions have been initiated, which impedes later oxidation reactions (Voziyan2002, Voziyan2005). On other hand, pyridoxamine also inhibits the oxidation of glycation products by scavenging free ROS, which neutralizes the reactive radicals and yields a less reactive PM- radical stabilized by resonance in the pyridine ring (Voziyan2005). The last inhibition route of AGE/ALE formation by PM consists in the reduction of reactive radical species by chelation of the active redox metal ions that catalyze their formation (Voziyan2005, Adrover2008, Ortega-Castro2009, Ortega-Castro2012) (Scheme 18).



## **2. Methodology**



## 2.1. The Schrödinger Equation

In quantum mechanics, the state of a system is described by its associated wave function  $\Psi$ . Given a molecular system of  $N$  electrons and  $M$  nuclei, its wave function and energy correspond to the solutions of the Schrödinger equation. According to the non-relativistic time-independent formulation, the Schrödinger equation adopts the form

$$H\Psi = E\Psi \quad [1]$$

where  $H$  is the Hamiltonian operator associated to the energy,  $E$ , of the molecular system. In atomic units, where the charge and mass of the electron,  $m$  and  $e$ , the reduced Planck constant,  $\hbar$ , and the Coulomb constant,  $1/4\pi\epsilon_0$ , are unity,  $H$  is

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{r_{ij}} - \frac{1}{2} \sum_{A=1}^M \sum_{B=1}^M \frac{Z_A Z_B}{R_{AB}} \quad [2]$$

being  $r_{iA}$  the distance between the  $i$ th electron and the  $A$ th nucleus,  $r_{ij}$  the distance between the  $i$ th and  $j$ th electrons,  $R_{AB}$  the distance between the  $A$ th and  $B$ th nuclei, and  $Z_A$  and  $M_A$  the charge and mass of the  $A$ th nucleus. The first two terms of [2] are the operators that take into account the kinetic energies of the electrons and the nuclei. The third operator corresponds to the Coulomb attraction between nuclei and electrons, while the fourth and fifth terms are the operators that correspond to the electron-electron and nuclei-nuclei Coulomb repulsions.

## 2.2. The Born-Oppenheimer Approximation

Since nuclei are significantly heavier than electrons, the latter move much faster than the nuclei. Thus a good approximation is to consider that electrons move around the fixed nuclei and that readapt instantaneously to their displacements. Accordingly, the kinetic energy of the nuclei, represented by the second term of [2], can be neglected. In addition, the last term of equation [2], which represents the nuclei-nuclei electrostatic repulsion, is constant since the nuclei are considered to be fixed. Therefore, the total molecular Hamiltonian operator is reduced to the electronic Hamiltonian,  $\hat{H}_e$ ,

$$\hat{H}_e = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{r_{ij}} - \frac{1}{2} \sum_{A=1}^M \sum_{B=1}^M \frac{Z_A Z_B}{R_{AB}}$$

Also, since the movement of nuclei and electrons are assumed to be decoupled, the total wave function of the system  $\Psi(\{\mathbf{r}_i, \mathbf{R}_A\})$  can be expressed as a product of a wave function describing the motion and interaction of electrons in the field of the fixed

nuclei,  $O_{el}(\{r_i\}, \{R_A\})$ , and a wave function describing the motion and interaction of the nuclei in the average field of the electrons,  $O_{nuc}(\{R_A\})$ .

$$O(\{r_i, R_A\}) = O_{el}(\{r_i\}, \{R_A\}) O_{nuc}(\{R_A\}) \quad [4]$$

The electronic part of the total wave function, also known as electronic wave function,  $O_{el}$ , is the eigenfunction of the electronic Hamiltonian given by [3] and the solution of the electronic Schrödinger equation

$$H H_{el} O_{el} = E_{el} O_{el} \quad [5]$$

The eigenvalues of the electronic Hamiltonian, which are also the solutions of the electronic Schrödinger equation, are the electronic energies,  $E_{el}$ . Accordingly, the total energy for a given set of nuclear coordinates is the summation of the electronic energy and the nuclear repulsion for such configuration

$$E_{tot} = E_{el} + \sum_{A=1}^M \sum_{B>1}^M \frac{Z_A Z_B}{R_{AB}} \quad [6]$$

Thus both the electronic energy and wave function depend *explicitly* on the electronic coordinates but *parametrically* on the nuclear coordinates. That is, the electronic Schrödinger equation has to be solved to obtain the specific  $O_{el}$  and  $E_{el}$  that adapt best to the given arrangement of the nuclei.

### 2.3. Pauli Exclusion Principle and Slater determinants

A complete description of the electrons in a molecular system requires that the electronic wave function fulfils the *antisymmetry principle*, which states that a many electron wave function must be antisymmetric with respect to the interchange of the coordinates (both space and spin) of any two electrons

$$O(x_1, \dots, x_i, x_j, \dots, x_j, x_i, \dots, x_N) = -O(x_1, \dots, x_j, x_i, \dots, x_i, x_j, \dots, x_N) \quad [7]$$

where  $x_i$  stands for the spatial plus the spin coordinates of the  $i$ th electron. Since the electronic Hamiltonian only depends on the spatial electron coordinates, the wave function solutions of the electronic Schrödinger equation,  $O_{el}$ , only depend on the spatial electron coordinates [4]. The spin coordinates,  $w$ , are intrinsic forms of angular momentum of, amongst other particles, fundamental particles such as electrons, which arise in the context of relativistic quantum mechanics. Therefore, the spin functions  $a(w)$  and  $P(w)$  are introduced *a posteriori* in the non-relativistic formulation to describe

the correct behaviour of electrons. It should be noted that the  $a(w)$  and  $B(w)$  spin functions are orthonormal.

$$\int dw a^*(w)a(w) = \int dw B^*(w)B(w) = 1 \quad [8]$$

$$\int dw a^*(w) B(w) = \int dw B^*(w) a(w) = 0 \quad [9]$$

The electronic Schrodinger equation [5] cannot be solved analytically for systems constituted by more than one electron. In practice, many electron wave functions are expressed in terms of many single electron wave functions. Such functions have to describe the spatial distribution and the spin state of the electron, and are known as spin orbitals,  $Z_i(d)$ , where  $x_i$  is the set of spatial plus spin coordinates. An easy way to build spin orbitals is by multiplying spatial orbitals,  $y_i(r)$ , by one of the two spin functions  $a(a)$  or  $B(a)$ .

When many electron wave functions are constructed, each electron is described by a spin orbital. However, not every combination of spin orbitals is valid due to the restriction given by [7]. A Slater determinant is a form of arranging the spin orbitals in such a way that all the electrons are described by all spin orbitals

$$\psi(x_1, x_2, \dots, x_N) = (N!)^{-1/2} \begin{vmatrix} \phi_1(x_1) & \phi_1(x_2) & \dots & \phi_1(x_N) \\ \phi_2(x_1) & \phi_2(x_2) & \dots & \phi_2(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(x_1) & \phi_N(x_2) & \dots & \phi_N(x_N) \end{vmatrix} = |Y\rangle \quad [10]$$

where  $(N!)^{-1/2}$  is a normalization constant. In [10] the row elements only contain the coordinates of a single electron and the column elements only contain a single spin orbital. The interchange of coordinates of two electrons, which involves the interchange of two rows, changes the sign of the determinant. In addition, a Slater determinant fulfils the Pauli Exclusion Principle since if two electrons are described by the same spin orbital, two columns will be equal and the wave function will be zero. Therefore, in a Slater determinant the motion of electrons with parallel spin is correlated, which is known as exchange correlation. However, a Slater determinant in which two columns only differ in the spin coordinate is different than zero, which involves that there is a probability greater than zero of finding two electrons simultaneously in the same region of space and that the motion of electrons with antiparallel spin is not correlated.

## 2.4. The Hartree-Fock Approximation

The Hartree-Fock method provides an approximate solution to the molecular wave function by considering a single Slater determinant.

$$\Phi \approx |Y_0\rangle \quad [11]$$

Given a set of  $K$  spin orbitals, the Hartree-Fock wave function of an  $N$  electron system is a determinant formed by the  $N$  spin orbitals with the lowest energies. In the following discussion, the lowest energy spin orbitals, also occupied spin orbitals, will be labelled by the indices  $a, b, c, \dots$  and the labels  $r, s, t, \dots$  will designate the remaining spin orbitals, known as virtual spin orbitals. Thus the Hartree-Fock method provides the best variational approximation to the ground state given by a single determinant. It should be noted that as any Slater determinant, the Hartree-Fock wave function fulfils the antisymmetry requirements but it does not describe the correlation in the motion of electrons with opposite spins. However, this approximation is central in quantum chemistry as it constitutes the starting point of more accurate approximations.

The application of the electronic Schrödinger equation [5] provides the solution of the Hartree-Fock wave function  $Y_0$  and the expectation value of the ground state energy,  $E_0$ , within the approximation.

$$H|\Phi_0\rangle = E_0|\Phi_0\rangle \quad [12]$$

$$\langle \Phi_0 | H | \Phi_0 \rangle = E_0 \quad [13]$$

Considering the form of the electronic Hamiltonian operator [3], equation [13] can be expressed as

$$E_0 = \langle \Phi_0 | H | \Phi_0 \rangle = \sum_a^N \langle h | a \rangle + \frac{1}{2} \sum_a^N \sum_b^N E_{ab} M_{ab} \quad [14]$$

where  $h$  is an operator,  $h$ , formed by the one electron kinetic energy and the Coulomb electron-nuclei attraction operators of [3]

$$h = \sum_{i=1}^N \frac{1}{2} \nabla^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}}$$

$\langle h | \rangle$  is the notation for one-electron integrals and  $\langle ij | kl \rangle$  designates the difference  $\langle ij | kl \rangle - \langle j | lk \rangle$ , where  $\langle ij | kl \rangle$  is a two-electron integral

$$\langle ij | kl \rangle = \int \psi_i^*(x) \psi_j^*(y) \psi_k(x) \psi_l(y) dx dy$$



$$(ij|kl) = (Z^i Z^j | X_k X_l) = \int dx^1 dx^2 \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_i^*(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \chi_k(\mathbf{r}_2) \chi_l^*(\mathbf{r}_2) / \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_i^*(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \chi_k(\mathbf{r}_2) \chi_l^*(\mathbf{r}_2) \quad [18]$$

The two-electron integrals of [14] can be separated in two types, namely  $(ab|ab)$  and  $(ab|ba)$ . The  $(ab|ab)$  integrals are the Coulomb integrals, which correspond to the classical electrostatic repulsion between the charge densities of electron 1,  $X_a^*(x_1)X_a(x_1)$ , and electron 2,  $X_b^*(x_2)X_b(x_2)$ .

$$(ab|ab) = \int d\mathbf{r}_1 d\mathbf{r}_2 X_a^*(\mathbf{r}_1) X_a(\mathbf{r}_1) X_b^*(\mathbf{r}_2) X_b(\mathbf{r}_2) \quad [18]$$

The  $\langle ab|ba \rangle$  integrals are the exchange integrals, which have no classical equivalency and arise from the Slater antisymmetrized product.

$$(ab|ba) = \int d\mathbf{r}_1 d\mathbf{r}_2 X_a^*(\mathbf{r}_1) X_b(\mathbf{r}_1) X_b^*(\mathbf{r}_2) X_a(\mathbf{r}_2) \quad [19]$$

When integrated over the spin coordinates, [18] and [19] are denoted by  $J^{ab}$  and  $K^{ab}$  respectively. The Coulomb operator  $J^b$  represents the mean repulsion on electron 1 caused by a second electron in the spin orbital  $X_b$

$$J_b^X a(\mathbf{r}_1) = \int d\mathbf{r}_2 X_b^*(\mathbf{r}_2) X_b(\mathbf{r}_2) X_a(\mathbf{r}_1) \quad [20]$$

And the Exchange operator  $K^b$  is defined by its effect when operating on a spin orbital  $X_a$

$$K_b^X a(\mathbf{r}_1) = \int d\mathbf{r}_2 X_b^*(\mathbf{r}_2) X_a(\mathbf{r}_2) X_b(\mathbf{r}_1) \quad [21]$$

According to [13], the energy  $E^0$  is a functional of the spin orbitals  $X_a$ . The Hartree-Fock equations arise from the minimization of  $E^0$  with respect to the set  $\{X_a\}$  under the restriction that the spin orbitals remain orthonormal

$$(a|b) = \int d\mathbf{r} X_a^*(\mathbf{r}) X_b(\mathbf{r}) = \delta_{ab} \quad [22]$$

Using the Lagrange method of undetermined multipliers and the constraint [22], the resulting functional of the spin orbitals is

$$L[\{X_a\}] = E^0[\{X_a\}] - \sum_{a,b} \lambda_{ab} ((a|b) - \delta_{ab}) \quad [23]$$

where the functional of the energy  $E^0[\{X_a\}]$  is given by [14] and  $\lambda_{ab}$  are the Lagrange multipliers.

Minimization of  $E_0$  is obtained by deriving  $L[\{\phi_a\}]$  and equalization to zero

$$5L = 2 \sum_a^N \{ \int d\mathbf{x} d\mathbf{x}_a (1) f^*(\mathbf{x}) X_a(1) + S(J - K(1)) X_a(1) - S \sum_b^N X_a^*(1) X_b(1) \} = 0 \quad [24]$$

Since the variation of the spin orbitals,  $\delta\phi_a(1)$ , is arbitrary and different than zero, the summation of the terms in brackets in [24] should be zero

$$i0 + \sum_b^N J(1) - K_b(1) X_a(1) = S \sum_b^N X_b(1) \quad a = 1, 2, \dots, N \quad [25]$$

resulting a system of  $N$  equations, which are the Hartree-Fock equations. The combination of operators in brackets of [25] is commonly known as the Fock operator, whose eigenfunctions are the spin orbitals and its eigenvalues are the orbital energies.

$$f = i(1) + \sum_b^N J(1) - K_b(1) \quad [26]$$

Eventually, the set of equations [25] can be expressed in the matrix form and after unitary transformations one obtains the canonical form of the Hartree-Fock equations

$$f \phi_a = S_a \epsilon_a \phi_a \quad [27]$$

The eigenfunctions of the canonical Hartree-Fock equations [27] are the canonical spin orbitals, which are delocalized over all the nuclei positions. For this reason are known as the canonical molecular spin orbitals. As shown in [26], the Fock operator depends on its eigenfunctions through the Coulomb [20] and Exchange operators [21]. As a result, the Hartree-Fock equations should be solved iteratively. Starting from a trial set of functions, the Fock operator is obtained and the eigenvalue problem is solved iteratively until convergence. For this reason, the Hartree-Fock method is also known as the Self-Consistent Field (SCF) method.

## 2.5. The Roothan-Hall equations

The inconvenience of [27] is that there is no analytical expression for the spin orbitals and numerical solution is required. Although numerical techniques are common in atomic calculations, there are no practical procedures for the numerical resolution of [27]. Roothan and Hall showed that the Hartree-Fock integro-differential equations can be transformed into a set of linear equations and solved with algebraic techniques.

In this method, which is also known as the Linear Combination of Atomic Orbitals or LCAO, a set of  $K$  known spatial basis functions  $\{\phi_i^M\}$  is introduced so that the molecular orbitals are expressed as linear combinations of the basis functions

$$\psi_i = \sum_K C_{Ki} \phi_K^M \quad i = 1, \dots, N \quad [28]$$

Then, the problem of finding  $N$  molecular orbitals of unknown mathematical expression becomes a problem of optimizing the coefficients of the linear expansions [28]. In the following discussion, the Roothan-Hall equations will be obtained for the restricted closed-shell Hartree-Fock method. In a restricted approximation, the  $\alpha$  and  $\beta$  spin orbitals are constructed from the same spatial orbitals. Additionally, the closed-shell systems are formed by an even number of electrons. So, in a restricted closed-shell method, the  $N/2$   $\alpha$  electrons are paired with the  $N/2$   $\beta$  electrons so that only  $N/2$  spatial orbitals are occupied. Once the spin is integrated, the Hartree-Fock equation [27] becomes

$$F \psi_i = \epsilon_i \psi_i \quad [29]$$

Substituting the molecular orbitals by the linear expansions given by [28], multiplying on the left by  $\langle \phi_j^M |$  and integrating, equation [29] becomes the matrix expression of the Roothan-Hall equations

$$FC = SCs \quad [30]$$

$C$  is a matrix whose  $K$  columns contain the  $K$  expansion coefficients of each orbital  $\psi$ ,  $s$  is a diagonal matrix of the orbital energies  $\epsilon_i$ ,  $F$  is the matrix representation of the Fock operator in the basis  $\{\phi_K^M\}$ ,  $S$  is the overlap matrix. The elements of the Fock and overlap matrixes,  $F_{ij}^{MV}$  and  $S_{ij}^{MV}$ , are respectively given by

$$S_{ij}^{MV} = \int \phi_i^M \phi_j^M d\tau = \langle \phi_i^M | \phi_j^M \rangle \quad [31]$$

$$F_{ij}^{MV} = \int \phi_i^M \left( \sum_{a=1}^{N/2} C_{a1} C_{a2} \left[ 2 \langle \phi_a^M | \phi_i^M \phi_j^M \rangle - \langle \phi_a^M | \phi_i^M \phi_j^M \rangle \right] \right) d\tau \quad [32]$$

where the notation  $\langle \phi_i^M | \phi_j^M \rangle$  is equivalent for spatial orbitals to the one-electron integrals  $\langle \phi_i^M | \phi_j^M \rangle$  [16], while  $\langle \phi_i^M \phi_j^M | \phi_k^M \phi_l^M \rangle$  and  $\langle \phi_i^M \phi_j^M | \phi_k^M \phi_l^M \rangle$  are respectively the two-electron Coulomb and Exchange integrals of the spatial orbitals analogous to  $\langle ij | kl \rangle$  [18] and  $\langle ij | kl \rangle$  [19].

From the expression of the Fock operator elements, a new matrix  $\mathbf{P}$  can be defined. Such matrix  $\mathbf{P}$  is known as density matrix and its  $UV$  element are given by the the product of the expansion coefficients of the  $U$  and  $V$  spatial orbitals,

$$P_{UV} = 2 \sum_{\mu} C_{\mu}^U C_{\mu}^V \quad [33]$$

which shows that the Roothaan-Hall equations are nonlinear and its solution requires an iterative procedure.

$$\mathbf{F}(\mathbf{C})\mathbf{C} = \mathbf{S}\mathbf{C}\mathbf{s} \quad [34]$$

Since the functions of the basis  $\{f_{\mu}\}$  are not orthogonal, the overlap matrix  $\mathbf{S}$  is not diagonal, in which case the matrix equation [34] would be reduced to

$$\mathbf{F}\mathbf{C} = \mathbf{C}\mathbf{s} \quad [35]$$

and its eigenfunctions and eigenvalues would be obtained by diagonalization of the  $\mathbf{F}$  matrix. The Lowdin transformation is one possible way of orthogonalizing the basis set  $\{f_{\mu}\}$ , in which the  $\mathbf{F}'$  and  $\mathbf{C}'$  matrixes are defined as

$$\mathbf{F}' = \mathbf{S}^{-1/2} \mathbf{F} \mathbf{S}^{-1/2} \quad [36]$$

$$\mathbf{C} = \mathbf{S}^{-1/2} \mathbf{C}' \quad [37]$$

The resulting transformed Roothaan-Hall equations that can be solved for  $\mathbf{C}'$  by diagonalization of  $\mathbf{F}'$  are

$$\mathbf{F}'\mathbf{C}' = \mathbf{C}'\mathbf{s} \quad [38]$$

The iterative solution of [38] provides the  $\mathbf{C}'$  matrix from which  $\mathbf{C}$  can be calculated by using [37] to define  $\mathbf{F}$  according to [33]. Then, the matrix equation [34] can be solved to obtain the proper eigenvectors  $\mathbf{C}$  and eigenvalues  $\mathbf{s}$ .

## 2.6. Form of the exact wave function and Electron Correlation

In the Hartree-Fock theory, the wave function is approximated by a single Slater determinant that is the best variational solution of the ground state. The resulting Hartree-Fock wave function describes correctly the correlated motion of electrons with parallel spin because of the properties of Slater determinants. However, the motion of electrons with opposite spin is not correlated. Although the Hartree-Fock wave function is a good initial approximation, the inclusion of electron correlation becomes necessary

for an accurate calculation of most atomic and molecular physical and chemical properties.

It can be easily demonstrated that the exact form of the wave function of an  $N$  electron system can be expressed as a linear combination of all the  $N$  possible Slater determinants  $|Y\rangle$  formed from a complete set of spin orbitals  $\{x\}$ . Furthermore, not only this approximation provides the exact wave function of the ground, but also all the wave functions of all possible excited states. Taking the Slater determinant of the ground state as a reference, the exact wave function expansion is

$$|\Psi\rangle = C_0|\Phi\rangle + \sum_{ra} S_{ra} |\Phi + V' a\rangle + \sum_{\substack{a<b \\ r<s}} S_{rs} |\Phi + V' a + V' b\rangle + \sum_{\substack{a<b<c \\ r<s<t}} S_{rst} |\Phi + V' a + V' b + V' c\rangle + \dots \quad [38]$$

The summations involving inequalities such as  $a<b$  stand for summations over all the occupied spin orbitals  $a$  and over all the occupied spin orbitals  $b$  higher in energy than  $a$ , and analogously for the summations over  $r<s$  in the case of virtual spin orbitals. Therefore [38] accounts for all the possible double, triple and higher excitations. Given that the employed basis set  $\{x\}$  is complete, the exact wave function is formed by the infinite possible  $N$ -electron determinants. Since each Slater determinant determines an electron configuration, this procedure is called Configuration Interaction (CI).

The eigenvalues of the Hamiltonian matrix associated to the exact wave function [38] expressed in the basis of the determinants  $|Y\rangle$  are the exact non-relativistic energies of the ground and excited states of the system. The correlation energy is defined as the difference between the exact ground state energy,  $E_0$ , and the Hartree-Fock-limit energy,  $E_{HF}(I)$ , which is obtained by using an infinite basis set

$$E_{corr} = E_0 - E_{HF}(I) \quad (I \rightarrow \infty)$$

In practice [38] can only be solved for finite basis sets in the so-called full CI approach. However, given a set of  $2K$  spin orbitals,  $\binom{2K}{N}$  possible determinants can be formed, which still is excessively large to be handled computationally for most systems.

The next logical approximation is to truncate [38] to a certain level of excitations. Alternatively, other theoretical methodologies improve the Hartree-Fock wave function to introduce the electron correlation effects by selecting a reduced number of elements of the full CI Hamiltonian matrix.

The most used approximations for the calculation of electron correlation are the Coupled Cluster (CC) and the Møller-Plesset (MP) methods, which use a Hartree-Fock Slater determinant as starting point. However, the Hartree-Fock wave function leads to incorrect results when the system is constituted by degenerate or quasi degenerate configurations. In these cases, several Slater determinants are required to describe the wave function, leading to multiconfigurational methods amongst which the most common are the Multiconfigurational Self Consistent Field (MCSCF), the Complete Active Site Self Consistent Field (CASSCF) and its perturbative version CASPT2.

## 2.7. Foundations of Density Functional Theory, the Hohenberg-Kohn Theorems

The Density Functional Theory, DFT, is an alternative formulation of the electronic problem in which the electron density, instead of the wave function, is the fundamental physical entity. DFT methods are interesting as they allow the inclusion of electron correlation in the solution of the Schrodinger equation at a low computational cost.

The rigorous fundaments of Density Functional Theory were set in two theorems proposed by Hohenberg and Kohn. The first one is known as the Existence Theorem and states that any observable of a stationary non-degenerate ground state can be calculated exactly from the electron density of the ground state. That is to say, any observable can be written as a functional of the electron density of the ground state.

This theorem is only valid for non-degenerate states and also requires that  $p(r)$  is a positive function defined in all the space and its integral over the entire space is equal to the number of electrons

$$p(r) > 0 \quad \text{and} \quad \int dr p(r) = N \quad [40]$$

The proof of the first theorem leads to a fundamental conclusion by which the ground state density  $p(r)$  is univocally connected to the external potential  $v(r)$ , i.e. the electron-nuclei potential. Furthermore, since both the number of electrons and their interaction potential with the nuclei are determined by  $p(r)$ , so are the Hamiltonian and the ground state wave function. Therefore, the electron density contains all the necessary information to obtain the expectation value of any observable of the ground state. The functional expression of the energy  $E[p]$  is of especial interest

$$E[p] = T[p] + V_{ne}[p] + V_{ee}[p] + (V_{NN}) \quad [41]$$

$T[p]$  and  $V_{ee}[p]$  are the functional of the electron kinetic energy and electron-electron repulsion potential, and are said to be universal functional since they do not depend on the external potential.  $V_{ne}[p]$  is the functional of the electron-nuclei attraction potential and  $V_{NN}$  is the internuclear repulsion term, which is added as a constant.

The second theorem or the Variational Theorem states that the electron density of a non-degenerate ground state can be calculated exactly by determining the density that minimizes the energy of the ground state. This theorem implies

$$E_0 < E_v[p] \quad [42]$$

where  $E_0$  is the exact energy of the ground state and  $E_v$  is the energy given by the external potential  $v(r)$  determined by the trial density  $p(r)$ . This theorem sets a

variational principle which allows the determination of the exact density of the ground state by calculating the density that minimizes the energy functional

$$\frac{\delta}{\delta \rho(r)} \left( E[\rho] - \mu \left( \int \rho(r) dr - N \right) \right) = 0 \quad [43]$$

However, this minimization requires that the N-representability, i.e. equation [42], is preserved. Thus the Lagrange method of undetermined multipliers with the constraint [42] can be considered to obtain the functional  $L[\rho]$

$$L[\rho] = E[\rho] - \mu \left( \int \rho(r) dr - N \right) \quad [44]$$

where  $\mu$  is a Lagrange multiplier and has the physical meaning of a chemical potential. The minima of the functionals  $L[\rho]$  and  $E[\rho]$  coincide since the second term of [44] is zero, which allows the minimization of  $E[\rho]$  by deriving  $L[\rho]$  and equalizing to zero

$$\frac{\delta}{\delta \rho(r)} \left( E[\rho] - \mu \left( \int \rho(r) dr - N \right) \right) = 0 \quad [45]$$

where the definition of a functional differential has been taken into account. The condition of minimum under the constrain is obtained when

$$\frac{\delta E[\rho]}{\delta \rho(r)} = \mu \quad [46]$$

Recovering the definition of  $E[\rho]$  given by [41], [46] converts to the Fundamental equation of Density Functional Theory

$$\frac{\delta E[\rho]}{\delta \rho(r)} = \mu = v_{\text{ext}}(r) + \frac{\delta E_{\text{kin}}[\rho]}{\delta \rho(r)} + \frac{\delta E_{\text{Hxc}}[\rho]}{\delta \rho(r)} \quad [47]$$

## 2.8. The Kohn-Sham method

The Hohenberg-Kohn theorems demonstrate that the electron density determines the Hamiltonian and the wave function of an N-electron system. This approach leads to wave function quantum mechanics so, to this point the Density Functional Theory does not involve any practical improvement with the respect to wave function-based methods. If the exact expression of  $E[\rho]$  was known, equation [47] could be solved with much less complexity than the wave function approaches.

Kohn and Sham proposed the use of a reference system of non-interacting electrons moving under a fictitious potential,  $v_s(r)$ , so that its electron density is the

same as that of the real system of interacting electrons. The advantage of this scheme is that, once the electron-electron interactions are neglected from the Hamiltonian, the Hartree-Fock method provides the exact solutions of non-interacting electron systems. In the reference system of non-interacting electrons, the energy can be expressed as

$$E[p] = T[p] + \int dr p(r) V_{ee}(r) \quad [48]$$

where  $T(p)$  is the kinetic energy of the electrons in the non interacting system. In the case of the real system, the energy can be partitioned in the following contributions

$$E[p] = T[p] + \int dr p(r) V_{ee}(r) + V_{ee}(p) \quad [49]$$

The term  $T(p)$  is the exact kinetic energy, which differs from  $T(p)$  of the reference system because in the real system the motion of each electron is affected (correlated) by the motion of the others. Adding and subtracting  $T(p)$  and  $J(p)$  to [49]

$$E[p] = T[p] + \int dr p(r) V_{ee}(r) + J[p] + (T[p] - T[p]) + (V_{ee}(p) - J[p]) \quad [50]$$

The introduced term  $J(p)$  is the total classical electron-electron Coulomb repulsion between the electrons that define the electron density  $p(r)$ .

$$J[p] = \frac{1}{2} \int dr_1 dr_2 p(r_1) p(r_2) \frac{1}{|r_1 - r_2|} \quad [51]$$

It should be noted that according to [51]  $J(p)$  takes into account the repulsion of all the electrons of  $p(r)$  with themselves so, there is an overestimation error of the repulsion energy caused by the self-interaction.

The term  $T(p)-T(p)$  of [50] is the correlation kinetic energy as it measures the difference between the kinetic energies of the correlated electrons in the real system and the uncorrelated electrons in the reference system. The term  $V_{ee}(p)-J(p)$  is the difference between the classical electron-electron interaction, including the self-interaction errors, and the exact quantum electron-electron interaction and it corresponds to the exchange-correlation energy. These two last terms of [50] are grouped as the total exchange-correlation energy

$$E_{xc} = (T[p] - T[p]) + (V_{ee}(p) - J[p]) \quad [52]$$

which leads to the expression for the energy functional

$$E[p] = T[p] + \int dr p(r) V_{ee}(r) + \int dr_1 dr_2 p(r_1) p(r_2) \frac{1}{|r_1 - r_2|} + E_{xc}(p) \quad [53]$$



Applying the Fundamental equation of DFT introduced in [49] to the energy functional [53] one obtains

$$M = \frac{ST_s[p]}{v_n(r)} + \int dr \frac{p(r)}{v_{xc}(r)} = \frac{8T_s[p]}{v_{eff}(r)} \quad [54]$$

Derivation of the energy expression for the reference system given by [49] shows analogue to that of the real system [53] with the exception of the explicit form of the potential. Therefore, the solution of [54] for the real system is the same than for the reference system of non-interacting electrons, which means that such solution can be obtained by a procedure analogue to the Hartree-Fock method. The resultant Kohn-Sham equations are the analogous to the Hartree-Fock equations

$$4S \quad X_i = \quad ,X_i \quad [55]$$

The orbitals of the set  $\{x\}$  are the Kohn-Sham orbitals and allow the calculation of the electron density from

$$P(r) = \left( \sum_i^N \right)^2 \quad [56]$$

The kohn-Sham Hamiltonian operator  $h_{KS}$  is defined by the one-electron kinetic energy operator and the effective potential  $v_{eff}(r)$

$$1 \quad [57]$$

$$V_{eff}(r) = v_n(r) + \int dr \frac{p(r)}{v_{xc}(r)} \quad [58]$$

Similarly to the Hartree-Fock procedure, starting from a set of approximate Kohn-Sham orbitals  $\{x\}$  the density is defined from [56], which in turn defines the potential  $v_{eff}(r)$  [58] and allows the resolution of [55] to obtain the new Kohn-Sham functions  $\{x'\}$ , which are used in a new iteration until the convergence is reached. In practice, the resolution of the Kohn-Sham equations is performed by using the respective Roothan-Hall equations (i.e. the Linear Combination of Atomic Orbitals approach), in which the Fock operator is substituted by the Kohn-Sham operator.

## 2.9. Exchange-correlation functional

The Kohn-Sham method provides a procedure for the minimization of the energy but does not solve the fact that the exact electron-electron repulsion functional,  $V_{ee}(p)$ , and the exact electron kinetic energy,  $T(p)$ , or alternatively the exchange-correlation potential,  $v_{xc}(r)$ , are not known. However, the potential of Density Functional Theory with respect to Hartree-Fock Theory is that it is possible to incorporate all the correlation energy providing that exchange-correlation potential,  $v_{xc}(r)$ , approaches the exact form.

From the total exchange-correlation energy,  $E_{xc}[p]$  [52], the exchange-correlation energy per particle,  $s_{xc}[p]$ , can be defined and related to the exchange-correlation potential  $v_{xc}(r)$

$$E_{xc}[p] = \int d\mathbf{r} p(\mathbf{r}) v_{xc}(\mathbf{r}) \quad [59]$$

$$v_{xc}(\mathbf{r}) = \frac{s_{xc}[p]}{p(\mathbf{r})} \quad [60]$$

where it is common to treat the exchange and correlation contributions of  $s_{xc}[p]$  separately

$$s_{xc}[p] = s_x[p] + s_c[p] \quad [61]$$

In the Local Density Approximation, LDA, the exchange contribution  $s_x[p]$  is calculated from the density considering that is constant as in a homogeneous gas of electrons and the correlation contribution  $s_c[p]$  is neglected or treated as a constant. The Local Spin Density Approximation or LSDA is the unrestricted version of the LDA approximation. In this approximation the exchange contribution is calculated separately for the  $\alpha$  and  $\beta$  electrons by taking into account only the densities generated by the  $\alpha$  and  $\beta$  electrons respectively. These approximations consider that the exchange-correlation effects are local and only depend on the density value at each point. Despite the simplicity of this approximation, LDA provides good results for solid-state calculations. However, the electron density of molecules is much more abrupt and irregular than in solids so, LDA does not describe correctly breaking and formation of chemical bonds.

The step to improve the local approximation consists on making the exchange and correlation contributions dependent on the density,  $p(r)$ , and also on the density gradients,  $\nabla p(r)$ . Accordingly, the density variation around each point is taken into account in addition to the density value. This correction, known as the Generalized Gradient Approximation or GGA, is considered to be semi-local instead of completely non-local. DFT calculations with GGA functionals improve LDA geometries and frequencies and average errors of several kcal/mol are obtained for thermochemical

quantities. However, the description of van der Waals interactions is not accurate due to the semi-local instead of completely non-local character of GGA functionals.

In the Meta-GGA approximation the exchange and correlation contributions depend on the density and density gradient, as in the GGA approximation, but also depend on the kinetic energy density  $i(r)$  and on the Laplacian of the density,  $\nabla^2 \rho(r)$ .

Since the HF method provides the exact exchange energy, a solution to improve the exchange contribution is to include part of the exact HF exchange energy. The adiabatic connection approach is used to calculate the amount of HF exchange energy introduced in the exchange functional and leads to the so-called hybrid functionals. Eventually, there is also the possibility of introducing HF exchange energy in the meta-GGA functionals, which constitutes the hybrid-meta-GGA approximation. The hybrid-GGA and especially the hybrid-meta-GGA functionals provide improved results with respect to GGA and meta-GGA functionals regarding the calculation of thermochemical and kinetic data and non-covalent interactions.

## 2.10. Continuum solvent models

Continuum solvent models can be considered the simplest approach to a liquid condensed-phase. In these models all the solvent molecules that solvate a given solute are removed and substituted by a dielectric continuum medium that is able to respond to the charge distribution of the solute with the purpose of representing the *average* solvent effects at thermal equilibrium. The basic quantity used for the definition and design of continuum solvent models is the free energy of solvation,  $\Delta G_{\text{solv}}$ .

The solvation energies available from experiment exhibit values varying from positive tens to negative hundreds of kcal/mol. The magnitude of solvation energies depends on the solute-solvent interaction, which can be partitioned in several contributions. The most important components are electrostatic interactions, cavitation and dispersion energies.

Unless the solute is apolar, electrostatic interaction energies are always attractive, i.e. stabilizing, and their existence involves that the solvent molecules show a response to the charge distribution of the solute, which is usually called reaction field. In turn, the response of the solvent also polarizes the solute until the stabilization by the mutual polarization balances the energy cost of polarization. In order to model the solvent-solute electrical response, an extra term should be included in the solute Hamiltonian operator.

The cavitation contribution to the solvation energy is not a solute-solvent interaction. Instead, it is related to the energy cost of creating the cavity that contains the solute in the dielectric continuum. So, cavitation energies are always positive or unfavourable to the solvation process. Dispersion energies take into account the favourable solute-solvent van der Waals interactions, in which all the non-classical electrostatic interactions are included.

There are various approximations to describe the cavities that contain the solute. The simplest cavity models are just a single sphere or an ellipsoid for the entire solute.

Such simple cavities show the advantage that the solute-solvent electrostatic interaction can be calculated analytically. A more realistic model is to use individual interlocking spheres on the atoms of the solute, which provide a solute-shaped cavity. In this approximation is typical to choose for each sphere the van der Waals radius multiplied by a constant, usually  $\sim 1.2$ . The resultant van der Waals surface obtained in this fashion may be sharp so, a refinement is to use a probe solvent molecule to define the solvent accessible surface, SAS. In practice, rough van der Waals surfaces because of their computational simplicity in front of SAS cavities. Alternatively, the solute electron density may be used to define its own cavity.

Cavities are important because of the cavitation energy term, but also because a common approximation is to consider that the solute-solvent dispersion interactions are proportional to the surface cavity. In addition, the electrostatic solute-solvent interactions are dependent on the definition of the cavities. There are several methods that provide solutions to the electrostatic problem such as the Poisson-Boltzmann methods or the Born/Onsager/Kirkwood models but in the case of quantum calculations, the Self-Consistent Reaction Field models are the most important.

The Poisson equation is a second-order differential equation that relates the electrostatic potential  $\langle j \rangle$ , the charge distribution  $\rho$ , and the dielectric constant  $\epsilon$

$$\nabla(\epsilon(r)\nabla\phi(r)) = -4\pi\rho(r) \quad [62]$$

In the case of a molecular shaped-cavity the Poisson equation has to be solved numerically. In order to do so, equation [62] is reformulated in terms of a surface integral over surface charges, which is numerically solved by dividing the surface in smaller fractions called tesserae. After the fragmentation of the surface, the total surface charge is each is also divided between the tesserae in surface charge elements  $o(r_i)$ . These surface charge elements are related to perpendicular component to the surface of the electric field  $\mathbf{F}$  generated by the solute on each surface element by

$$4\pi s o(r_i) = (\epsilon - 1)\mathbf{F}(r_i) \quad [63]$$

Once the surface charges  $o(r_i)$  are determined, the electrostatic potential  $\phi(r)$  is calculated as

$$\phi(r) = \int \frac{o(r_i)}{|r - r_i|} dr_i \quad [64]$$

Finally, the introduction of the surface potential  $\phi(r)$  into the solute Hamiltonian operator causes the polarization of the solute wave function. However, the surface potential is also determined by the electric field generated at the cavity surface by the solute charge distribution [63], which involves that  $\phi(r)$  has to be solved iteratively in the so-called Self-Consistent Reaction Field, SCRF.

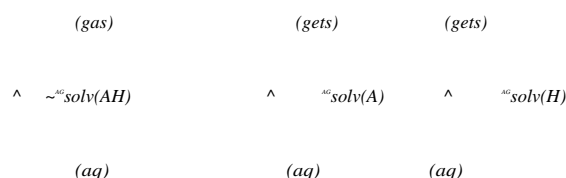
## 2.11. Computational determination of $pK_a$ values

Theoretical  $pK_a$  values can be obtained by considering thermodynamic cycles, which can also require the use of experimental values. Three of the most used cycles will be introduced in the following section. In all the thermodynamic cycles, the free energy of a reaction that involves the deprotonation of a studied acid species AH is calculated in gas-phase,  $AG_{gas}$ . Additionally, in all cases the free energies of solvation and desolvation of the involved species are calculated from gas-phase and liquid phase calculations,  $AAG_{sol}$ . Eventually the aqueous deprotonation free energy is calculated as

$$AG_a = AG_{gas} + AAG_{sol} \quad [65]$$

These terms adopt different forms depending on each thermodynamic cycle. The first thermodynamic cycle is constituted by three thermodynamic processes, namely the desolvation of the acid AH, its deprotonation in gas-phase and, finally, the solvation of its conjugate base A and the proton H+ (Scheme 19)

### Cycle 1



Scheme 19. Thermodynamic cycle which considers the dissociation of the acid species in the conjugate base and an isolated proton. The total charge of the acid and the conjugate base are represented by  $q$  and  $q-1$  respectively.

According to Scheme 19, the free energy of the reaction in gas-phase,  $AG_{gas}$ , is the deprotonation free energy of AH and is calculated as

$$AG_{gas} = G_{gas}(H+) + G_{gas}(A^{q-1}) - G_{gas}(AH^q) \quad [66]$$

Since the electronic energy of the proton is zero, its gas-phase enthalpy,  $H_{gas}(H+)$ , is obtained by adding up the translational energy ( $E=3/2RT$ ) and  $PV=RT$ , which at 298 K is 1.48 kcal/mol. The gas-phase entropy,  $\mathcal{E}_{gas}(H+)$ , is calculated by considering the Sackur-Tetrode equation for gas phase monoatomic species (McQuarrie1970), which makes  $TS_{gas}(H+)=-7.76$  kcal/mol at 298 K and 1atm. Eventually, the sum of  $H_{gas}(H+)$  and  $T\mathcal{E}_{gas}(H+)$  makes  $G_{gas}(H+)=-6.28$  kcal/mol.

The  $pK_a$  determination considers a 1 M concentration standard state for the deprotonation free energy in solution. Aqueous phase calculations also use a standard state of 1M but the gas-phase deprotonation free energies are calculated for a standard state of 1atm. Therefore, the gas-phase free energies,  $AG_{gas}$ , must be referred to 1M by taking into account the factor  $RT\ln 24.46$

$$AG_{gas} (1M) = AG_{gas} (1atm) + RT \ln 24.46 \quad [67]$$

The calculation of the solvation free energy increment,  $\Delta AG_{solv}$ , in Cycle 1 requires the value of  $AG_{solv}(H^+)$ . This magnitude is taken from experiment but there is not a unique value in the current literature. Two of the most used values are -264.0 kcal/mol (Palascak2004) and -265.9 kcal/mol (Tissandier1998), both referred to a standard state of 1M in the gas phase and 1M in the aqueous phase.

$$\Delta AG_{solv} = AG_{solv} (H^+) + AG_{solv} (A^{q-1}) - AG^{\circ} (AH^q) \quad [68]$$

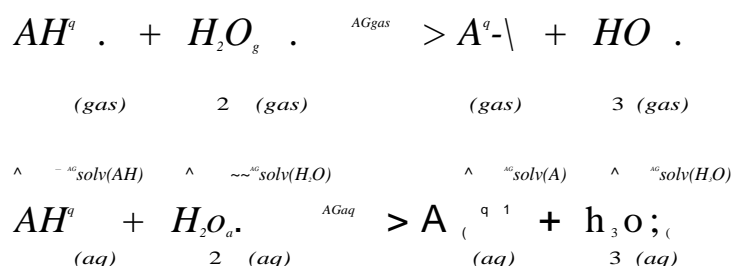
Eventually, the  $pK_a$  for this thermodynamic cycle is obtained from

$$pK_a = \frac{\Delta AG}{RT \ln 10} \quad [69]$$

Since the uncertainty in  $AG_{solv} (H^+)$  can reach several kcal/mol (Palascak2004) and an error of 1.4 kcal/mol in  $AG_{aq}$  yields an error of 1  $pK_a$  unit in the predictions, it is advisable to propose an alternative for the treatment of the isolated proton in  $pK_a$  calculations.

Accordingly, in the second thermodynamic cycle, the proton is substituted by the hydronium cation and a water molecule is added in the reactants side to balance the chemical equation (Scheme 20).

### Cycle 2



Scheme 20. Thermodynamic Cycle 2 which considers an acid base reaction between an acidic species and water. The total charges of the acid and the conjugate base are represented by  $q$  and  $q-1$  respectively.

The values of  $G_{gas}(H_2O)$ ,  $\Delta G_{solv}(H_2O)$ ,  $G_{gas}(H_3O^+)$  and  $\Delta G_{solv}(H_3O^+)$  are available from experiment in the literature (Palascak2004, Pliego2000) or alternatively can be obtained from theoretical calculations. In this cycle,  $AG_{gas}(1atm)$  equals  $AG_{gas}(1M)$  because there is the same number of moles in both sides of the chemical equation. The  $pK_a$  values according to this cycle are obtained as follows

$$AG_{gas} = G_{gas} (H_2O) + G_{gas} (A^{q-1}) - G_{gas} (AH^q) - G_{gas} (H_2O) \quad [70]$$

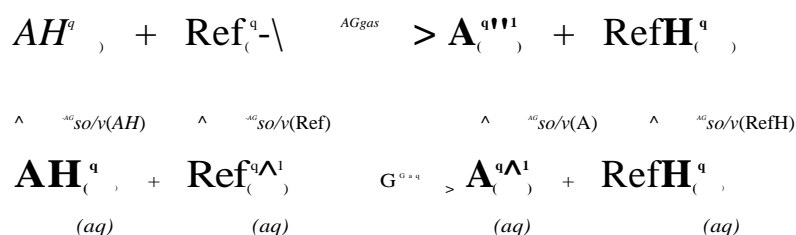
$$AAG_{sol} = AG_{sol}(H_3O^+) + AG_{sol}(A^{q-1}) - AG_{sol}(AH^q) - AG_{sol}(f_2O) \quad [71]$$

$$P^a_{aH} = \frac{AG}{RT \ln 10} \frac{f_a}{f_2O} - \log[H_2O] \quad [72]$$

According to [71], the experimental value reported for the free energy of solvation of the hydronium cation,  $AG_{sol}(H_3O^+) = -110.4$  kcal/mol (Pliego2000), is used in combination with that reported for water,  $AG_{sol}(H_2O) = -6.32$  kcal/mol (Palascak2004). Alternatively, theoretical calculations can be done to obtain these values with the purpose of reducing empiricism in the  $pK_a$  determinations.

The third thermodynamic cycle is based on an *isodesmic* reaction, in which the type of bonds broken in the reactants are the same as those formed in the products (Cramer2004) (Scheme 21).

### Cycle 3



Scheme 21. Thermodynamic Cycle 3 which considers an acid base reaction between an acidic species AH and a reference species RefH. The total charge of the acids and the conjugate bases are represented by  $q$  and  $q-1$  respectively.

The main advantage of this approximation in  $pK_a$  calculations is the expected cancellation of errors between the solvation free energies of the charged species in the reactants and products sides. Moreover, the experimental free solvation free energies of the proton and the hydronium cation are not required in this cycle. However, this approximation can be subjective as long as a reference species, RefH, should be chosen, if possible, possessing similar geometry, electrostatic distribution and acidity than AH.

In cycle 3, as in cycle 2,  $AG_{gas}(1atm)$  equals  $AG_{gas}(1M)$ . In this case, the calculated  $AG_{aq}$  cannot be directly related to the  $pK_a$  of the studied acid AH as it is referred to the  $pK_a$  difference with the reference species

$$AG_{gas} = G_{gas}(RefH^q) + G_{gas}(A^{q-1}) - G_{gas}(AH^q) - G_{gas}(RefH^{q-1}) \quad [73]$$

$$AAG_{gas} = AG_{gas}(RefH^q) + AG_{gas}(A^{q-1}) - AG_{gas}(AH^q) - AG_{gas}(RefH^{q-1}) \quad [74]$$

$$P^a_{aH} = \frac{AG_{aq}}{RT \ln 10} - pK_a(RefH)$$

1751

## 2.12 The CBS-QB3 method

The CBS-QB3 method belongs to the so-called Complete Basis Set Model Chemistry methods originally developed by Peterson and co-workers (Nyden1981, Petersson1983, Petersson1988, Petersson1991, Montgomery1994, Ochterski1996, Montgomery1999, Montgomery2000). The accurate calculation of atomic and molecular energies requires the convergence of both the single particle (basis set) and the  $n$ -particle energies (order of perturbation or CI). Since most of the energy contributions arise from low perturbation levels, CBS models involve low-level (SCF and ZPE) calculations on large basis sets, mid-sized basis sets for second-order corrections and small basis sets for high-level corrections. They include an extrapolation in order to correct the Møller-Plesset second-order energies to the complete basis set. Additionally, empirical and spin contamination corrections are included. Specifically, the CBS-QB3 method involves the following steps:

- (I) B3LYP/6-311G(2d,d,p) geometry optimization
- (II) B3LYP/6-311G(2d,d,p) frequencies with a 0.99 scale factor for the ZPE
- (III) UMP2/6-311+G(3d2f,2df,2p) energy and CBS extrapolation
- (IV) MP4SDQ/6-31+G(d(f),p) energy
- (V) CCSD(T)/6-31G<sup>†</sup> energy

Eventually, the electronic CBS-QB3 energies are calculated as

$$E_{\text{CBS-QB3}} = E_{\text{MP2}} + E_{\text{MP4}} + E_{\text{CCSD(T)}} + E_{\text{CBS}} + E_{\text{emp}} + E_{\text{int}} \quad (74)$$

where  $AE_{\text{CBS}}$  is the term correcting the basis set truncation error in the second-order energies, and the energy terms  $E_{\text{MP4}}$ ,  $E_{\text{CCSD(T)}}$ ,  $E_{\text{emp}}$  and  $E_{\text{int}}$  are calculated from the following equations

$$E_{\text{MP4}} = E_{\text{MP4(SDQ)/6-31+G(d(f),p)}} - E_{\text{MP4(SDQ)/6-31+G(d(f),p)}} \quad (75)$$

$$E_{\text{CCSD(T)}} = E_{\text{CCSD(T)/6-31+Gt}} - E_{\text{MP4(SDQ)/6-31+Gt}} \quad (76)$$

$$E_{\text{Emp}} = -0.00579 \sum_{i=1}^p \sum_{M=1}^f \left( \frac{N_{\text{virt}}+1}{M} \right)^{\frac{1}{N_{\text{virt}}+1}} \quad (79)$$

$$E_{\text{int}} = -0.00954[(S^2) - S_i(S_i - 1)] \quad [80]$$



### 2.13. Molecular Dynamics Simulations

Temperature effects on the nuclear motion can be satisfactorily approximated in gas-phase by several approximations such as the rigid rotor and the harmonic oscillator. However, in the case of liquid phase and solutions, the elevated number of degrees of freedom and possible configurations involves that the partition functions which connect to the macroscopic observables should be estimated from a limited sampling of the phase space. The molecular dynamics simulation techniques generate series of time-correlated points in the phase space which are known as trajectories. Given the initial coordinates and velocities of the particles that constitute the system, Newton's second law of motion, which in its differential form is

$$-\frac{dV}{d\mathbf{r}} = m \frac{d^2 \mathbf{r}}{dt^2} \quad [81]$$

where  $V$  is the potential energy at position  $\mathbf{r}$  and the  $\mathbf{r}$  is a vector of the coordinates of all particles. This equation can be used to predict the future positions and velocities of all the particles in the system. In practice, the integration of [81] has to be numerical for small time increments  $\Delta t$ , also called time steps, which in the case of atomic simulations are of the order of  $10^{-16}$ - $10^{-15}$  s.

There exist several integration algorithms for solving Newton's equations. In the Verlet algorithm, given a set of initial coordinates  $\mathbf{r}_i$  at time  $t$  the positions at time  $t + \Delta t$ , later than the initial time, can be approximated by a Taylor expansion

$$\mathbf{r}_{i,t+\Delta t} = \mathbf{r}_i + \mathbf{v}_i \Delta t + \frac{1}{2} \mathbf{a}_i (\Delta t)^2 + \frac{1}{6} \mathbf{b}_i (\Delta t)^3 + \dots \quad [82]$$

$$\mathbf{r}_{i,t+\Delta t} = \mathbf{r}_i + \mathbf{v}_i (\Delta t) + \frac{1}{2} \mathbf{a}_i (\Delta t)^2 + \frac{1}{6} \mathbf{b}_i (\Delta t)^3 + \dots \quad [83]$$

being  $\mathbf{v}_i$ ,  $\mathbf{a}_i$ ; and  $\mathbf{b}_i$ ; vectors formed by the velocities, accelerations, hyperacceleration of each particle. The positions at a time step earlier than the initial time are also given by [83] if  $\Delta t$  is substituted by  $-\Delta t$

$$\mathbf{r}_{i,t-\Delta t} = \mathbf{r}_i - \mathbf{v}_i (\Delta t) + \frac{1}{2} \mathbf{a}_i (\Delta t)^2 - \frac{1}{6} \mathbf{b}_i (\Delta t)^3 + \dots \quad [84]$$

Adding [83] and [84] allows the determination of the positions  $\mathbf{r}_{i,t+\Delta t}$  and  $\mathbf{r}_{i,t-\Delta t}$  at times  $t + \Delta t$  and  $t - \Delta t$

$$\mathbf{r}_{i,t+\Delta t} = (2\mathbf{r}_i - \mathbf{r}_{i,t-\Delta t}) + \mathbf{a}_i (\Delta t)^2 \quad [85]$$

where the accelerations of each particle have to be calculated from Newton's second law of motion [81] providing the potential acting on each particle is known and the forces can be calculated. In order to generate the trajectories, the integration procedure and the forces have to be solved successively.

Different types of molecular dynamics simulations are distinguished depending on the form of the potential  $V$  that governs the motion of the nuclei according to [81]. If  $V$  is chosen to be defined by a molecular mechanics force field, then the simulations are known as *classical* molecular dynamics simulations. However, the potential  $V$  can be determined from a quantum operator as in the Born-Oppenheimer approximation [3]. In such case, the potential and its derived forces will be of quantum nature since they are solved from the time-independent Schrodinger equation. In fact, the nuclei will "move" on a Born-Oppenheimer potential energy surface, thus the simulation will be considered as *ab initio* despite of the fact that the nuclei follow the classical Newton's equations of motion.

#### 2.14. Metadynamics simulations: Free Energy calculations and study of rare events

In practice, in a standard molecular dynamics simulation, only a limited subset of configurations is usually sampled due to computational limitations. Currently, the simulation length when using *ab initio* simulation techniques is typically in the range of tens to hundreds of picoseconds. However, the timescale of many relevant physical and chemical events such as chemical reactions or changes in the secondary structure of proteins is considerably larger. Therefore, as mentioned, it is not possible to sample such phenomena with standard molecular dynamics simulations.

The metadynamics technique (Laio2002) allows the sampling of events whose timescales are longer than the typical simulation lengths by forcing the system to explore new configuration space. In order to keep track of the sampled configurations and to avoid that the system revisits them, a small number  $n$  of collective variables (CV)  $O(x)$  are chosen, which are defined as a function of the system coordinates  $x$ .

$$a = (O_1(x), \dots, O_n(x)) \quad [86]$$

Once the metadynamics simulation starts, a series of Gaussian-shaped potentials are dropped with period  $T$ , which builds a history-dependent bias potential  $V(O(x), t)$  that is associated to the selected collective variables. At a time  $t$  of the metadynamics simulation, the history potential is

$$V(\sigma(x), t) = W \sum_{k: t_k < t} \exp \left[ - \sum_{i=1}^n \frac{(O_i(x) - O_i(t_k))^2}{2\sigma_i^2} \right] \quad [87]$$

where  $W$  is the height of each Gaussian in the potential units,  $ba_i$  is the width of the Gaussian for the  $i$ th collective variable,  $a_i$  is the value of the  $i$ th collective variable at time  $t$  and the  $a_i$  values are the values of the  $i$ th collective variable recorded at times  $kx < t$ .

The inclusion of the history-dependent potential [87] adds new forces to those given by the intrinsic classical or ab initio potential  $f(x)$ , whose magnitude depends on the values of the collective variables at each simulation time

$$f(t)' = f(t) - \left( \frac{\partial V(\sigma(x), t)}{\partial \sigma} \right) \left( \frac{\partial \sigma(x)}{\partial x} \right) \quad [88]$$

In the early stages of the metadynamics simulation, the evolution of the system will be mostly trapped in its intrinsic free energy surface. However, as the metadynamics progresses, the addition of Gaussian potentials compensates the forces of the free energy surface and the system is able to overcome the free energy barriers associated to the collective variables. Accordingly, it can be shown that if the Gaussian parameters  $W$ ,  $ha$  and the deposition rate parameters are well chosen, the history-dependent potential approaches the negative value of the free energy potential of the system

$$\lim_{t \rightarrow \infty} V(\sigma(x), t) = -F(\sigma) \quad [89]$$

Thus a correct metadynamics simulation provides an estimation of the Free energy surface associated to the chosen collective variables, allowing the study of rare events in affordable simulation lengths.

It should be noted that the choice of the collective variables is crucial in a metadynamics simulation and some aspects are to be considered (Barducci2011), namely: the number of collective variables should be as low as possible to keep the convergence of the free energy surface within a reasonable simulation length; the chosen collective variable should provide a clear distinction of the different free energy wells; and finally, all the slow modes that are relevant to the studied event should be included as collective variables.



## **3. Objectives**



Pyridoxal, pyridoxamine and their Schiff bases formed with amino acids have been extensively studied concerning the acid base chemistry of the heteroatoms. However, the experimental studies of the carbon acidities in the Schiff bases are restricted to the most abundant protonation states and tautomers in aqueous solution.

Proton transfer reactions involving C $\alpha$  and C4' are ubiquitous in the PLP-catalyzed reactions of amino acids in solution and in enzymes. Despite of this, there is only one experimental work concerning the carbon acidity of a PLP Schiff base in an enzymatic medium. In addition, this study is restricted to a single protonation state for the external aldimine due to the architecture of the active site, and the carbon acidity was roughly estimated with an uncertainty of 8 kcal/mol, which corresponds to an uncertainty of 5  $pK_a$  units. Additionally, there are no studies addressing the carbon acidity of C4' atom in the PMP Schiff bases, which is essential to understand how reaction specificity is achieved in each PLP-catalyzed reaction.

Different computational strategies have been successfully used for accurate  $pK_a$  determinations of a broad variety of compounds such as carboxylic acids, alcohols, amines, phosphoranes and phenols, amongst others. Several computational approaches previously reported in the literature together with new computational strategies will be tested to evaluate the accuracy on the predicted  $pK_a$  values of carbon acids relevant to amino acid chemistry and PLP-catalyzed reactions. Additionally, the carbon acidity of C4' in the Schiff bases of PMP will also be calculated with the aim to shed light on the mechanisms by which reaction specificity achieved in PLP-dependent enzymes.

The carbon acidities of C $\alpha$  and C4' will also be calculated in the Schiff bases that chelate metal ions in aqueous solution. The  $pK_a$  values calculated by computational methods correspond to the microscopic dissociation processes. Therefore, comparison between the calculated acidities of the complexed and free Schiff bases will help clarify whether the apparent Schiff base activation is due to real ligand activation or to an increase in the concentration of the Schiff bases by stabilization in the chelate form. Eventually, our computations also aim to provide a plausible explanation for the absence of metal ions in PLP-dependent enzymes.

In order to make the most of the computational strategies used for the prediction of  $pK_a$  values, analogous strategies will be adapted and evaluated for the prediction of stability constants ( $\log\beta$  values) of metal complexes. This is one of the current challenges of computational chemistry as the most accurate  $\log\beta$  values show uncertainties of 6-8 kcal/mol.

Apart from thermodynamic calculations, kinetic information on the deprotonation and reprotonation of the C $\alpha$  carbon in the Schiff bases is also important to completely understand the origins of catalysis by PLP. Accordingly, the activation barriers for the deprotonation and reprotonation reactions of the C $\alpha$  in the PLP Schiff bases by several acid/base catalysts will be calculated computationally. This study will provide important information about the lifetimes of the carbanionic intermediates and their evolution in the active sites that present different acid catalyst residues.

The last section of this dissertation will focus on the carbon-carbon bond formation and cleavage of amino acids catalysed by PLP. In particular, the aldol formation/retro aldol cleavage will be studied in relation to the carbonyl scavenging properties of PLP Schiff bases in aqueous solution. On one hand, this study is aimed to understand the PLP and PMP inhibition mechanisms of glycation reactions of biomolecules in the cell. On other hand, the results of this study will be combined with the information obtained for the kinetic stability of the carbanions against reprotonation. This comparison aims to provide information of the partially understood retro aldol mechanism in Serine hydroxymethyltransferase and related PLP-dependent enzymes.

Given the lack of consensus in relation with the PLP-catalyzed decarboxylation reactions, QM and QM/MM metadynamics simulations together with QM and ONIOM *static* calculations will be carried out to gain insight in the origins of catalysis by PLP and the role of the intramolecular proton transfer involving the imine and phenol groups. For this purpose, the decarboxylation of ornithine catalysed by PLP will be studied in gas phase, aqueous solution and in the active site of Ornithine decarboxylase.







## **4. Results**



The Results chapter is divided in three sections, namely assessment of computational protocols for the prediction of  $pK_a$  values (Results 4.1.), studies on the C-H acidity of Ca and C4' carbons of PLP and PMP Schiff bases (Results 4.2.) and studies on Ca-C bond breaking and formation of PLP Schiff bases (Results 4.3.).



#### 4.1. Evaluation of computational strategies for the calculation of $pK_a$ values and $\log P$ values of metal complexes

- 4.1.1. Simplification of the CBS-QB3 method for predicting gas-phase deprotonation free energies

R. Casasnovas, J. Frau, J. Ortega-Castro, A. Salva, J. Donoso, F. Muñoz *Int. J. Quantum Chem.*, **2010**, 110, 323-330

- 4.1.2. Absolute and relative  $pK_a$  calculations of mono and diprotic pyridines by quantum methods

R. Casasnovas, J. Frau, J. Ortega-Castro, A. Salva, J. Donoso, F. Muñoz *J. Mol. Struct. THEOCHEM*, **2009**, 912, 5-12

- 4.1.3. Avoiding gas-phase calculations in theoretical  $pK_a$  predictions

R. Casasnovas, D. Fernández, J. Ortega-Castro, J. Frau, J. Donoso, F. Muñoz *Theor. Chem. Acc.*, **2011**, 130, 1-13

- 4.1.4. Theoretical calculations of stability constants and  $pK_a$  values of metal complexes in solution: application to pyridoxamine-copper(II) complexes and their biological implications in AGE inhibition

R. Casasnovas, J. Ortega-Castro, J. Donoso, J. Frau, F. Muñoz *Phys. Chem. Chem. Phys.*, **2013**, 15, 16303-16313





4.1.1. Simplification of the CBS-QB3 method for predicting gas-phase deprotonation free energies

Pages 63-70 are not displayed in this version of the present dissertation due to copyright restrictions. The information is accessible with the following reference:

"Simplification of the CBS-QB3 method for predicting gas-phase deprotonation free energies" R. Casanovas, J. Frau, J. Ortega-Castro, A. Salvà, J. Donoso, F. Muñoz *Int. J. Quantum Chem.*, **2010**, *110*, 323-330

<http://dx.doi.org/10.1002/qua.22170>

4.1.2. Absolute and relative  $pK_a$  calculations of mono and diprotic pyridines by quantum methods

Pages 73-80 are not displayed in this version of the present dissertation due to copyright restrictions. The information is accessible with the following reference:

"Absolute and relative  $pK_a$  calculations of mono and diprotic pyridines by quantum methods" R. Casanovas, J. Frau, J. Ortega-Castro, A. Salva, J. Donoso, F. Muñoz *J. Mol. Struct. THEOCHEM*, **2009**, 912, 5-12

<http://dx.doi.org/10.1016/j.theochem.2008.11.020>

---

#### 4.1.3. Avoiding gas-phase calculations in theoretical $pK_a$ predictions

Pages 83-95 are not displayed in this version of the present dissertation due to copyright restrictions. The information is accessible with the following reference:

"Avoiding gas-phase calculations in theoretical  $pK_a$  predictions" R. Casasnovas, D. Fernández, J. Ortega-Castro, J. Frau, J. Donoso, F. Muñoz *Theor. Chem. Acc.*, **2011**, *130*, 1-13

<http://dx.doi.org/10.1007/s00214-011-0945-5>



4.1.4. Theoretical calculations of stability constants and  $pK_a$  values of metal complexes in solution: application to pyridoxamine-copper(II) complexes and their biological implications in AGE inhibition







## 4.2. Studies on the carbon acidities of PLP and PMP Schiff bases

- 4.2.1. Theoretical study on the distribution of atomic charges in the Schiff bases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms

R. Casasnovas, A. Salvà, J. Frau, J. Donoso, F. Muñoz *Chem. Phys.*, **2009**, 355, 149-156

- 4.2.2. C-H Activation in Pyridoxal-5'-phosphate Schiff Bases: The Role of the Imine Nitrogen. A Combined Experimental and Computational Study

R. Casasnovas, M. Adrover, J. Ortega-Castro, J. Frau, J. Donoso, F. Muñoz *J. Phys. Chem. B*, **2012**, 116, 10665-10675

- 4.2.3. C-H Activation in Pyridoxal-5'-phosphate and Pyridoxamine-5'-phosphate Schiff Bases: Effect of Metal Chelation. A Computational Study

R. Casasnovas, J. Frau, J. Ortega-Castro, J. Donoso, F. Muñoz, *J. Phys. Chem. B*, **2012**, 117, 2339-2347



4.2.1. Theoretical study on the distribution of atomic charges in the Schiff bases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms

Pages 115-122 are not displayed in the online version of the present dissertation due to copyright restrictions. The information is accessible with the following reference:

"Theoretical study on the distribution of atomic charges in the Schiff bases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms" R. Casasnovas, A. Salvà, J. Frau, J. Donoso, F. Muñoz *Chem. Phys.*, **2009**, 355, 149-156

<http://dx.doi.org/10.1016/i.chemphys.2008.12.006>

4.2.2. C-H Activation in Pyridoxal-5'-phosphate Schiff Bases: The Role of the Imine Nitrogen. A Combined Experimental and Computational Study

Pages 125-135 are not displayed in the online version of the present dissertation due to copyright restrictions. The information is accessible with the following reference:

"C-H Activation in Pyridoxal-5'-phosphate Schiff Bases: The Role of the Imine Nitrogen. A Combined Experimental and Computational Study" R. Casasnovas, M. Adrover, J. Ortega-Castro, J. Frau, J. Donoso, F. Muñoz *J. Phys. Chem. B*, **2012**, *116*, 10665-10675

<http://dx.doi.org/10.1021/jp303678n>





4.2.3 C-H Activation in Pyridoxal-5'-phosphate and Pyridoxamine-5'-phosphate Schiff Bases: Effect of Metal Chelation. A Computational Study

Pages 139-147 are not displayed in the online version of the present dissertation due to copyright restrictions. The information is accessible with the following reference:

"C-H Activation in Pyridoxal-5'-phosphate and Pyridoxamine-5'-phosphate Schiff Bases: Effect of Metal Chelation. A Computational Study" R. Casasnovas, J. Frau, J. Ortega-Castro, J. Donoso, F. Muñoz, *J. Phys. Chem. B*, **2012**, 117, 2339-2347

<http://dx.doi.org/10.1021/ip311861p>



### **4.3. Studies on the carbon-carbon bond breaking and formation of PLP Schiff bases**

4.3.1. Non-enzymatic Pyridoxal 5'-Phosphate-catalyzed aldol condensation between amino acids and sugars. An inhibition mechanism of Advanced Glycation End-Products (AGEs) formation

4.3.2. Extraordinaire decarboxylation rates catalyzed by modestly efficient enzymes. A QM/MM metadynamics study on the enzymatic and nonenzymatic pyridoxal 5'-phosphate-catalyzed decarboxylation of amino acids



4.3.1. Non-enzymatic Pyridoxal 5'-Phosphate-catalyzed aldol condensation between amino acids and sugars. An inhibition mechanism of Advanced Glycation End-Products (AGEs) formation





## **Non-enzymatic Pyridoxal 5'-Phosphate-catalyzed aldol condensation between amino acids and sugars. An inhibition mechanism of Advanced Glycation End-Products (AGEs) formation**

Rodrigo Casasnovas, Juan Frau, Joaquín Ortega-Castro, Josefa Donoso, Francisco Muñoz

*Departament de Química. Universitat de les Illes Balears. E-07122 Palma de Mallorca, Spain*

*e-mail. [juan.frau@uib.es](mailto:juan.frau@uib.es)*

### **Abstract**

A non-enzymatic pyridoxal 5'-phosphate (PLP)-catalyzed aldol condensation reaction between amino acid carbanions and carbonyl compounds has been studied with Density Functional Theory calculations as an inhibition mechanism of Advanced Glycation End-products (AGEs) formation. Schiff base formation between pyridoxal 5'-phosphate and amino acids enhances the acidity of the alpha carbons adjacent to the imine group to  $pK_a$  values of 12-17 in water solution. The calculations show that, once the carbanion is formed in aqueous solution, the free energy activation barrier for the carbon-carbon bond formation is only 4.6 kcal/mol in the worst-case scenario. However, the water-catalyzed reprotonation of Ca requires a much larger activation free energy (i.e. 15.8 kcal/mol) whereas only the carboxylic acid-catalyzed neutralization of the carbanion shows a lower free energy activation barrier (i.e. 3.7 kcal/mol). Therefore, the carbonyl scavenging reaction takes place preferentially to the carbanion neutralization under physiological pH conditions. Additionally, the hydrolysis yields a thermodynamically stable reduced carbonyl adduct and recovers the PLP, which is able to act as catalyst in subsequent reactions. These computations, which are in agreement with reported experimental results on related PLP-catalyzed reactions, suggest that carbonyl scavenging by PLP Schiff bases is an efficient mechanism of inhibition of AGEs formation and of prevention of their related medical conditions. Concerning the catalysis of PLP-dependent enzymes, our results support that the role of Glu53 in Serine hydroxymethyltransferase is to protonate the Schiff base carbanion at Ca to avoid the protonation at C4' by other residues and, also, to avoid the reversion of the retro-aldol reaction.

## 1. Introduction

Non-enzymatic glycation reactions occur in the physiological environment between reducing sugars or reactive carbonyl species (RCS) and amino groups of proteins, aminophospholipids and nucleic acids (1-4). The resulting product from the condensation between amino groups and sugars is an imine or Schiff base adduct. The Schiff base may then undergo an isomerization reaction yielding a 1-imine-1-ketose species known as Amadori compound. Schiff bases and Amadori compounds facilitate the degradation of those biomolecules on which are formed because of their susceptibility to oxidation reactions (1-4).

Hydroxyl, hydroperoxyl species (i.e. OH<sup>-</sup> and HOO<sup>-</sup>) and related compounds (i.e. RO<sup>-</sup> and ROO<sup>-</sup>) are powerful oxidizing species known as reactive oxygen species (ROS) which cause degradation of glycated biomolecules under physiological conditions. In addition, an excess of transition metals in the organism, such as Cu(II) and Fe(III), promotes these oxidation reactions. Furthermore, the resulting products of Schiff base and Amadori compound oxidations are new RCS and ROS species which further propagate uncontrolled glycation and degradation of biomolecules and tissues (5, 6) (Scheme 1).

The final degradation products of Schiff bases and Amadori compounds, also known as advanced glycoxilation and lipoxidation end-products or commonly AGEs and ALEs, are compounds of diverse chemical nature. Glycation of biomolecules and especially formation of AGEs and ALEs have been related to ageing (7-12), diabetes (7-10), atherosclerosis (7-10), tissue degradation (7-12), inflammatory diseases (7, 8, 12) and neurodegenerative illnesses such as Alzheimer's and Parkinson's diseases (13, 14). Therefore, the proposal and study of inhibition mechanisms of AGEs and ALEs formation is of great interest.

Vitamin B6, particularly in the form of pyridoxamine (PM), exhibits high inhibitory activity in several routes of AGE/ALE formation. Recent studies performed in our group (15-18) as well as by Voziyan et. al (19, 20) showed that PM is highly reactive towards carbonyl groups of sugars and RCS for the formation of stable Schiff bases, which avoids glycation reactions of free amino groups of biomolecules. Moreover, PM also forms stable Schiff bases with the carbonyl groups of Amadori compounds, hindering their subsequent oxidation reactions (15-17, 19, 20). Another inhibition route consists in the neutralization of ROS (radical scavenging), which protects glycation products from degradation (20). Additionally, pyridoxamine forms stable chelates with transition metals, diffculting their catalytic effect on oxidation reactions of biomolecules (20-24) (Scheme 1).

Pyridoxal-5'-phosphate (PLP) is another B6 vitamer that also inhibits glycation reactions by protecting the free amino groups of biomolecules in the form of Schiff base adducts (25). However, the most recognized biologic role of PLP is as electrophilic catalyst in the metabolism of amino acids (26, 27). Formation of Schiff bases between PLP and amino acids activates the C $\alpha$  carbon for decarboxylation, racemization, transamination and retro-aldol reactions. Pyridoxamine, specifically pyridoxamine-5'-phosphate (PMP), also participates in the second half of transamination reactions.

Recent experimental and theoretical works show that Schiff base formation between glycine and PLP reduces the  $pK_a$  of the Ca atom from 29 for the zwitterionic species in solution at neutral pH to  $pK_a \sim 17$  (28-30) (Scheme 2A). Computational studies also showed that C4' atom of pyridoxamine Schiff bases exhibits  $pK_a \sim 12-13$  in aqueous solution (Scheme 2A) and that the acidity of both Ca and C4' in the Schiff bases of PLP and PMP can be highly increased at neutral pH by metal chelation (30, 31). These data point out that a small, although non negligible, fraction of Schiff base is present in the carbanionic form under physiological conditions.

Interestingly, Richard and co-workers (28, 29, 32) reported that Schiff base carbanions of glycine and 5'-deoxypyridoxal (DPL) do not incorporate deuterium to any carbon atoms as a consequence of reprotonation by deuterated water. Instead, the products resulting from an aldol condensation between carbanionic Schiff bases and a carbonyl group of a second DPL molecule were characterized (Scheme 2B). Dalling et al. (33) reported that a mixture of PLP and alanine initially forms a Schiff base which yields pyruvate as the transamination product and a pyruvate dimer ( $O_2CC(OH)(CH_3)CH_2COCO_2^-$ ) as by-product. These investigations highlight the nucleophilic power of the Schiff base carbanions and their affinity for carbonyl groups. Furthermore, it has been suggested that aldol addition reactions between PLP Schiff bases and carbonyl groups should occur in physiological conditions (32).

In this work, we propose a new inhibition mechanism of AGEs and ALEs formation which consists in the reduction of sugars and reactive carbonyl species (RCS) by aldol addition of PLP/PMP Schiff base carbanions. Particularly, this study focuses on the condensation between the common carbanion of PLP-glycine and PMP-glyoxylate with the simplest aldose, glyceraldehyde. Density Functional Theory calculations were carried out to calculate the free energy profiles of three possible reaction routes. It is worth noting that neutralization of the Schiff base carbanion species is a reaction which competes with the aldol condensation and, therefore, affects the extension of carbonyl trapping. Therefore, protonation reactions by water and several representative species of acid catalysts in physiological conditions were also calculated. The activation free energies were confronted in order to determine whether carbonyl scavenging is a probable inhibition mechanism of AGEs/ALEs formation.

## 2. Computational Details

Density Functional Theory (DFT) calculations were performed with the Gaussian 09 software (34). Solvent effects were introduced with the SMD solvent model (35) in all calculations. The geometries of all reactants, intermediates and products were optimized with the M06-2X (36) functional and the 6-31+G(d) basis set. The same functional and basis set were employed to characterize all the optimized structures either as energy minima by the absence of imaginary frequencies or as transition state structures by the presence of a unique imaginary frequency in which the displacement of the nuclei is concordant with the reaction. Thermodynamic information at 298.15 K (i.e. thermal and entropic contributions to free energies) was also extracted

from the vibrational calculations. The considered standard state is 1 atm and 1 mol/l. Intrinsic Reaction Coordinate (IRC) calculations were carried out on each transition state to confirm that they connect the desired energy minima through the minimum energy path.

Refined energies of every structure were obtained by carrying out calculations at the M06-2X/6-311++G(2df,2pd) level of theory, also in the presence of the SMD solvent model. The refined free energies were obtained by adding up the M06-2X/6-311++G(2df,2pd) potential energies (calculated at 0 K) with the thermal and entropic corrections obtained in the vibrational analyses.

According to the standard Transition State Theory, equation 1 was used to obtain the microscopic kinetic constants  $k$  of each elementary step from the corresponding activation free energy ( $\Delta G^\ddagger$ ). In eq 1,  $T$ ,  $k_B$ ,  $h$  and  $R$  stand for the absolute temperature, Boltzmann, Planck and the gas constants, respectively.

### 3. Results and Discussion

#### 3.1. Acid-base catalyzed protonation/deprotonation of Schiff bases

Once the PLP and PMP Schiff bases are formed with amino acids and keto acids respectively, the acidity of Ca and C4' carbons are greatly enhanced (28-32). As a result, either carbon can be deprotonated generating small amounts of a carbanion species in water solution (28-30, 32). The rate of reprotonation of this carbanion limits the aldol condensation with carbonyl groups, so it is necessary to determine the activation free energy barriers of such protonation reactions.

Deprotonation of Ca in the PLP-glycine Schiff bases and C4' in the PMP-glyoxylate Schiff bases yield the same carbanion species. As reported in previous works, Ca is more basic than C4' (30, 31) so, the activation energy barrier of protonation at Ca should be lower than that of C4'. Therefore, the protonation barriers at Ca set the upper bound for the activation barriers of aldol addition. That is, for the carbonyl scavenging reaction to be significant, its activation free energy barrier has to be lower to those of carbanion protonation reactions.

In addition to water as protonation agent, which corresponds to protonation by solvent, acetic acid-, methylammonium-, phenol-, methanol- and methanethiol-catalyzed reactions have been studied. Such species represent the acidic groups of protein residue sidechains (i.e. aspartic and glutamic acids, lysine, tyrosine, serine and threonine and cysteine respectively). Dihydrogen phosphate-catalyzed protonation is also considered since this species is present as buffer agent together with hydrogen phosphate anion in the physiological medium.

The calculated activation free energies and reaction free energies for the protonation of Ca at 298.15 K are reported in Table 1. The difference between the

activation free energies ( $AG^*$ ) and the activation potential energies at 0 K including zero-point energy corrections ( $AE^*$ , Table 1) varies between 0.1 and 1.2 kcal/mol. Similarly, the reaction free energies and potential energies (i. e.  $AG_r$  and  $AE_r$ ) differ by 0.1-1.1 kcal/mol. These differences point out that the studied proton transfer reactions are essentially governed by stereoelectronic factors, whereas thermal and entropic contributions have minor influence. These results are in line with recent computational and experimental results of Griswold et al. (38), which point out that hyperconjugation has a significant contribution to the Ca-H bond destabilization in PLP Schiff bases. The torsion values of H-Ca-N<sub>im</sub>-C4' dihedral in the transition state structures show that proton transfer reactions take place almost perpendicularly to the plane formed by the pyridine ring and imine moiety (Table 1), which supports the hypothesis formulated by Dunathan (39). In this way, the negative charge of the carbanion species remains delocalized across the  $\pi$  system until the formation of the Ca-H bond. Additionally, these proton transfers are shown to be strongly directional in the transition state structures as deduced from the fact that the sum of Ca-H and X-H nearly coincides with the Ca-X distance (Table 1), being X the heteratom of the acid catalyst that donates the proton to Ca in each case.

The fastest proton transfer, i.e. the one which presents the lowest activation free energy (3.7 kcal/mol), corresponds to acetic acid-catalyzed reaction, which is the most acid species amongst the considered catalysts. Contrarily, methanol and water solvent are the least acid considered species and, accordingly, show the highest activation energies ( $AG^* = 15.8$  kcal/mol and  $AG^* = 15.7$  kcal/mol) (Table 1, Figure 1).

Inspection of the free energy values reported in Table 1 and Figure 1 for the reactions catalyzed by dihydrogen phosphate, phenol, methanethiol and methylammonium show that the acidity (i.e.  $pK_a$ ) is not the only determining factor of the activation and reaction energies. It should be noted that the calculated reaction free energies,  $AG_r$ , do not strictly correspond to the thermodynamic reaction free energies but to the free energy differences between the reactants and products complexes. Therefore, in addition to the  $pK_a$  difference between the proton donor and acceptor, these free energies also depend on the intermolecular interactions and solvent stabilization of the reactant and product complexes.

Dihydrogen phosphate, phenol, methanethiol and methylammonium have intermediate  $pK_a$  values with respect to acetic acid and water/methanol. Accordingly, their catalyzed protonation reactions exhibit intermediate activation free energies (Table 1, Scheme 3). The lowest activation energy in this group corresponds to the reaction catalyzed by  $CH_3NH_3^+$ , which is the most basic species ( $pK_a = 10.66$ ,  $AG^* = 5.9$  kcal/mol). However, methylammonium is the only cationic acid catalyst, so the interaction with the Schiff base carbanion is more effective, especially considering that the carboxylate group of the amino acid is next to the Ca carbon (Figure 2A).

Oppositely, dihydrogen phosphate is the only anionic acid but the activation free energy associated to its catalyzed proton transfer is 8.5 kcal/mol, which is lower than those of the reactions catalyzed by neutral phenol and methanethiol (Table 1, Figure 1). However,  $H_2PO_4^-$  is 2.8-3.1  $pK_a$  units more acidic than phenol and methanethiol. In addition,  $H_2PO_4^-$  establishes a hydrogen bond with the Ca-carboxylate group which

decreases the electrostatic repulsion and places the other proton in a favorable position over the Ca carbon to favor the proton transfer reaction (Figure 2B).

The importance of the global charge in the acid catalysis is also exemplified by comparing the methanethiol and methylammonium catalyzed reactions. The acidities of these species are very similar (i.e.  $pK_a$  10.66 and 10.33 for  $CH_3SH$  and  $CH_3NH_3^+$  respectively) but the activation free energy of the protonation catalyzed by methanethiol is 4 kcal/mol higher (Table 1, Figure 1). On the other hand, both phenol and methanethiol are neutral species, have similar acidities and the activation free energies of their catalyzed reactions only differ by 1.1 kcal/mol (Table 1).

From the relative disposition of the aromatic rings of phenol and the Schiff base in the optimized structures, these species establish n-n interactions in the reactants, transition state and products complexes (Figure 2C). The average of least distances between the atoms in the two aromatic systems is 3.42 Å, 3.37 Å and 3.28 Å respectively for the optimized reactants, transition state and products complexes. The free energy of the protonation reaction ( $\Delta G$ ) with phenol is 4.7 kcal/mol more exergonic than that with methanethiol but similar to the reaction free energy of the reaction with dihydrogen phosphate (where an extra hydrogen bond is formed), which highlights the contribution to the complex stability of the n-n interaction.

The microscopic kinetic constants of each catalyzed protonation reaction were calculated by introducing the theoretical activation free energies in eq 1 (Table 1). As mentioned previously, the thermal and entropic contributions to the free energies are significantly lower than the stereoelectronic effects. Therefore, the standard Transition State Theory approach is appropriate for the calculation of the kinetic constants. It is worth noting that nuclear quantum effects were not considered despite being important for reactions involving light nuclei such as hydrogen. Consequently, the values in Table 1 are a lower bound to the quantum-corrected kinetic constants.

### 3.2. Carbonyl scavenging

The proposed reaction can be divided in four steps, namely: A) Aldol condensation between the Schiff base carbanion and glyceraldehyde (structures 1-4); B) nucleophilic addition of  $H_2O$  at C4' and formation of a carbinolamine species (structures 4-6); C) protonation of the amino group at C4' (structures 6-10); D) release of the final products and recovery of the PLP catalyst (structures 10-12). Different reaction routes starting from the same reactants that yield the same final products were investigated but equivalent structures in each route have been equally labeled (Scheme 3, Scheme 4 and Scheme 5). Table 2 reports the free energies of each transition state and intermediate structures at 298.15 K, which have been referred to the initial reactant complex 1. These values are represented in Figure 3 as the free energy profiles of each reaction route.

Previous computational works have studied the reaction mechanisms of transamination between PLP Schiff bases and amines (40, 41) and Schiff base formation between PLP and amino acids (42, 43) or PMP and carbonyl compounds (18). These reactions were studied in gas-phase, aqueous solution or in an enzyme active site, and

show common mechanistic features with the reaction studied in this work. Transimination starts with the displacement of the imine nitrogen due to the nucleophilic attack of a second amine group, which generates a gem-diamine intermediate. Subsequently, the gem-diamine is hydrolyzed yielding the Schiff base formed with the attacking amine group and releasing the amine species that formed the former Schiff base. On the other hand, during the Schiff base formation reaction, the nucleophile amine attacks a carbonyl group generating a carbinolamine species. The later dehydration of the hemiaminal carbon yields the Schiff base. Precisely, Schiff base formation is the opposite reaction to the hydrolysis studied in this work, which takes place after the aldol condensation.

### 3.2.1. Nucleophile addition to the carbonyl group, aldol condensation

The aldol condensation, which has also been called Claisen-type addition by previous authors (28, 29, 32), starts with the nucleophile attack of the Ca of the Schiff base carbanion to the carbonyl group of glyceraldehyde. The Schiff base carbanion exhibits all the atoms of the carboxylate, imine and pyridine moieties in the same plane. Such disposition stabilizes the carbanion because of the delocalization of the negative charge from the Ca carbon throughout the entire  $\pi$ -system. Furthermore, this conformation makes possible the formation of two intramolecular hydrogen bonds between the carboxylate and phenoxide groups with the protonated imine nitrogen.

In the reactant complex 1, the carbonyl group of glyceraldehyde lies parallel to the molecular plane of the Schiff base carbanion. In this complex, there are two explicit water molecules which favor such arrangement by forming hydrogen bonds simultaneously between the carboxylate and phenoxide groups with the carbonyl oxygen. Additionally, an alcohol group of glyceraldehyde forms another hydrogen bond with the remaining carboxylic oxygen. These interactions place the carbonyl carbon of glyceraldehyde at a close distance of Ca (2.96 Å) and in a favorable orientation for the condensation to take place (Scheme 3, Scheme 4, Scheme 5).

The condensation reaction is identical for routes A and B, the Ca carbon is added to the carbonyl group of glyceraldehyde via a transition state 3-TS to yield a Schiff base 4 (Scheme 3, Scheme 4). The structural changes from reactants 1 to the transition state 3-TS are minor. All the hydrogen bond interactions in 1 are retained in 3-TS with small distance variations, and the only significant change is the distance reduction between Ca and the carbonyl carbon (labeled as Cp) down to 2.27 Å. The forming bond occurs almost perpendicularly to the  $\pi$ -system of the Schiff base (i.e. Cp-Ca-N<sub>im</sub>-C4' torsion is 81.2°) analogously to protonation/deprotonation reactions of Ca (Figure 4A).

As a consequence of the preorganization of the reactants in complex 1, the activation free energy for this reaction is only 4.6 kcal/mol (Table 2, Figure 3), which according to eq 1 corresponds to a microscopic kinetic constant of  $2.4 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$  for this step (Table 2). Considering only the condensation step, this value shows that carbonyl scavenging is kinetically favorable with respect to reprotonation of the carbanion by water solvent and by the rest acid catalysts with the exception of acetic acid. This result

is in agreement with the experiments of Richard and co-workers performed at neutral pH (28, 32), because acetic acid or carboxylic acids in general are completely deprotonated in such conditions.

For routes A and B, the thermodynamics of this first step is also favorable since the free energy of the reaction is -7.1 kcal/mol (Table 2, Figure 3). The resulting product 4 is a Schiff base of PLP and a modified glycine amino acid. The Cot-Cp bond distance is 1.56 Å and the original intramolecular hydrogen bonds of the Schiff base carbanion reactant complex are maintained in 4. From the reactant complex 1 to product 4, the major changes in the  $\pi$ -system are observed in the Ca-N<sub>im</sub> bond distance, which increases from 1.33 Å to 1.45 Å, whereas the remaining bond distances (i.e. Nim-C4', C4'-C4 and those involving the pyridine ring atoms) oscillate between 0.01 and 0.05 Å. The hydrogen bond between an explicit water and oxygen O3' is broken in product 4 so that the formed alkoxide oxygen in Cp is completely solvated by such water molecule.

In route C (Scheme 5), the 5'-phosphate group indirectly participates in the condensation reaction.

As mentioned, the Schiff base carbanion in solution exhibits a planar geometry in the carboxylate, imine and pyridine atoms. The phosphate group at carbon 5' also lies in the same plane to minimize the electrostatic repulsion between the negative oxygens and the  $\pi$  electrons. However, once the complex 1 is formed, the dihedrals C4-C5-C5'-O5' and C5-C5'-O5'-P can rotate so that the phosphate group moves close to the glyceraldehyde molecule forming the new complex 2. In this conformation, the water molecule that bridges the O3' and the carbonyl oxygen atoms substitutes the hydrogen bond formed with the O3' atom to establish a new hydrogen bond with an oxygen atom of the phosphate group, which stabilizes the reactant complex by 2.2 kcal/mol (Table 2, Figure 3). These changes do not modify the remaining hydrogen bonds of complex 1 and the distance between Ca and carbonyl carbon atoms Cp is 2.89 Å and the Cp-Ca-Nim-C4' torsion is 82.4°.

The condensation reaction in route C occurs via a transition state 3-TS that is analogous to that of routes A and B except for the hydrogen bond involving the phosphate group. The Ca-Cp bond distance is 2.30 Å in 3-TS and the approach direction of glyceraldehyde is also perpendicular to the  $\pi$ -system (i.e. Ca-Cp-Nim-C4' torsion is 81.2°) (Figure 4B). The activation free energy of this step is 6.1 kcal/mol from complex 2, therefore being 1.5 kcal/mol higher than the activation free energy in routes A and B. However, since complex 1 is also the starting point of route C, the global activation barrier is 3.9 kcal/mol (Figure 3) and the overall microscopic kinetic constant is  $6.7 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$  (Table 2).

Similarly to routes A and B, the condensation reaction in route C is thermodynamically favorable since the resulting product 4 is 7.6 kcal/mol more stable than complex 2 (9.8 kcal/mol with respect to the initial complex 1) (Table 2, Scheme 4). In fact, the resulting product 4 of route C is the same species as that of routes A and B but for the conformation of the phosphate group which allows the water molecule solvating the alkoxide oxygen at Cp to keep a second hydrogen bond with the phosphate group.



## 3.2.2. Formation of the carbinolamine compound

The formed Schiff base 4 can be hydrolyzed yielding a modified amino acid and regenerating PLP. After the aldol condensation, the next step in the reaction is the first step of the hydrolysis and consists in the nucleophilic addition of a water molecule to carbon C4' to produce a hemiaminal species also known as carbinolamine.

For routes A and B, the water molecule solvating the alkoxide anion is the closest to carbon C4' (i.e. 3.15 Å measured from the water oxygen). The nucleophilic addition is simultaneous to a proton donation from the water molecule to the alkoxide oxygen via a concerted transition state 5-TS. The remaining water molecule in 5-TS establishes hydrogen bonds with the nucleophile water oxygen and the phenoxide oxygen O3'. From the Schiff base 4 to the transition state 5-TS, the water oxygen  $O_{w1}$  distance to C4' decreases to 2.14 Å (Figure 5A). Simultaneously, the water  $O_{w1}$ -H bond increases from 1.02 Å to 1.55 Å and the hydrogen bond with the alkoxide oxygen decreases from 1.60 Å to 1.02 Å. These bond changes point out that although the water addition and the proton transfer take place in a unique step, the reaction is not symmetric because the proton is almost entirely transferred to the alkoxide oxygen in the transition state structure (Figure 5A). As observed for the protonation/deprotonation and carbonyl condensation reactions at Ca, the direction of the nucleophile attack to C4' is perpendicular to the plane formed by the imine and pyridine moieties (i.e.  $O_{w1}$ -C4'-C4-C3 torsion is 95.8°) (Figure 3A).

In reaction route C, the nucleophile water molecule that solvates the alkoxide group also forms a second hydrogen bond with one oxygen atom of the phosphate group (Figure 5B). Such water molecule is at 3.41 Å of C4' in the Schiff base 4 of route C, somewhat farther than in routes A and B. However, this distance is reduced down to 2.22 Å in the transition state 5-TS of route C, resembling that of routes A and B (i.e. 2.14 Å). In fact, the water addition to C4' in route C is also accompanied by proton transfer from the water nucleophile to the alkoxide oxygen at Cp (Figure 5B). Moreover, the remaining water molecule forms hydrogen bonds with the attacking water and the phenoxide oxygen O3' as in routes A and B. From 4 to 5-TS, the water  $O_{w1}$ -H distance lengthens from 1.00 Å to 1.65 Å while the distance between this hydrogen and the oxygen alkoxide decreases from 1.61 Å to 1.00 Å (Figure 5B). Therefore, the proton transfer is again completed before the transition state, as without the participation of the phosphate group.

A study of the Intrinsic Reaction Coordinate (IRC), which corresponds to the minimum energy path that connects reactants and products with the transition state, reveals that the first half-reaction consists in the proton transfer to the oxygen alkoxide for all routes A, B and C. In this way, a hydroxide anion is generated in situ, enhancing the nucleophilic power of the water oxygen atom for the attack on C 4'. Once this species is formed, the  $O_{w1}$ -C4' bond formation in the second half-reaction takes place downhill in the potential energy surface according to the IRC calculations. Therefore, the free energy barriers of this step (i.e. 7.7 kcal/mol and 9.8 kcal/mol for routes A/B and C respectively) can be associated to the energetic cost of forming the reactive hydroxide anion species.

A previous study on the Schiff base formation between PLP and methylamine in water solution, carried out in our group (42), shows that the oxygen carbon ( $O_{w1}$ -C4') distance in the transition state is considerably shorter (i.e. 1.907 Å). Furthermore, the activation free energy for the water addition to C4' is 18.7 kcal/mol, much larger than that of the present reaction. It should be noted that in the standard Schiff base formation reaction, the attacking water donates a proton to the phenoxide oxygen O3' (42) instead of to an alkoxide anion, as in the current case. Likewise, Oliveira et al. (43) reported similar results for the formation of the internal aldimine (i.e. the Schiff base formed between the PLP cofactor and the  $\epsilon$ -amino group of a lysine residue of an enzyme) in the active site of Ornithine decarboxylase. According to their results, the required free energy of activation is 15.9 kcal/mol for the water addition to C4', and the  $O_{w1}$ -C4' distance in the transition state is 1.89 Å. Again, this is a standard Schiff base formation, so the nucleophile water donates a proton to the phenoxide oxygen O3', which is much less basic than the alkoxide anion in the current reaction. A comparison of these data with our results clearly shows that the largest contribution to the activation energy cost is the deprotonation of the nucleophile water molecule and that such process is catalyzed by the presence of the alkoxide anion at Cp.

The resulting intermediate 6 is a heminaminal or carbinolamine compound at C4' carbon. As for the Schiff base 4, the formed carbinolamine 6 is the same compound for all three reaction routes A, B and C, so the free energy differences shown in Table 2 and Figure 3 result from different conformations and hydrogen bonding patterns of the carbinolamine 6.

The free energy of reaction for route B is -6.6 kcal/mol and the resulting product 6 shows the same hydrogen bond distribution than in the transition state 5-TS. The formed alcohol group at C4' establishes a hydrogen bond with the alcohol at Cp (1.96 Å) while the amine at C4' donates two hydrogen bonds to the phenoxide O3' (2.19 Å) and carboxylate (2.22 Å) oxygen atoms (the interactions are not shown in Scheme 4 for clarity purposes). Finally, the remaining water molecule bridges the alcohol at C4' (2.09 Å) and the phenoxide oxygen (1.73 Å) (Scheme 4).

In route A, the carbinolamine undergoes a conformational change, so the amine at C4' lies over the pyridine ring (i.e. 68.2° for the dihedral Nim-C4'-C4-C3). In the new conformation, the hydrogen bond with the phenoxide oxygen O3' remains at 2.20 Å but the one formed with the carboxylate group is virtually broken. Additionally, in this conformation, the alcohol group of C4' is neither hydrogen bonded with the alcohol of Cp. The peculiar feature of this conformation is that the water molecule acts as hydrogen bond acceptor of the alcohol group at C4' (2.03 Å) and as hydrogen bond donor with the amine group also at C4' (1.89 Å). Overall, the reaction free energy of this step in route A is -3.6 kcal/mol, which is approximately 3.0 kcal/mol less than that of route B (Table 2, Figure 3).

In route C, the conformation of carbinolamine 6 is almost identical to that of route A except for an additional hydrogen bond with one oxygen atom of the 5'-phosphate group at 1.76 Å. However, this additional hydrogen bond makes the resulting carbinolamine 6 more stable than those of routes A and B, and the reaction free energy for this step in route C is -4.4 kcal/mol (Table 2, Figure 3).

## 3.2.3. Formation of zwitterionic carbinolamine

Once the carbinolamine 6 is formed, the next step in the hydrolysis process is the protonation of the amino group at C4', which favors its elimination. Different mechanisms for the proton transfer between the geminal groups of C4' have been examined in previous works, which include the participation of: solvent molecules (40, 41), the phenoxide oxygen O3' (18, 42) or specific residues in the case of some enzymatic reactions (43). In the present work we have considered the solvent-mediated reaction in route A, the analogous solvent-mediated reaction with indirect participation of the 5'-phosphate group in route C and the intramolecular two-step proton transfer via oxygen O3' in route B.

In the carbinolamine 6 of route A, the hydroxyl group at C4' donates a hydrogen bond to the water molecule which, in turn, donates a hydrogen bond to the amine at C4'. So as to produce the zwitterionic carbinolamine 10, the water molecule donates the proton to the amine nitrogen and captures the proton from the alcohol in a concerted step. The transition state structure 7-TS together with the Intrinsic Reaction Coordinate study show that the two proton transfer reactions do not occur simultaneously. In the first half-reaction, the proton transfer to the amine nitrogen is almost complete (i.e.  $d_{O_{w2}-H} = 1.52 \text{ \AA}$  and  $d_{H-N_{C4'}} = 1.11 \text{ \AA}$  in the transition state 7-TS) while the proton transfer from the hydroxyl group to the water oxygen has barely started (i.e.  $d_{O_{C4'}-H} = 1.01 \text{ \AA}$  and  $d_{H-O_{w2}} = 1.67 \text{ \AA}$ ) (Figure 6A).

The carbinolamine 6 compounds of routes A and C only differ from each other in the presence of an extra hydrogen bond with the phosphate group in route C. Consequently, the formation of the zwitterionic carbinolamine 10 in route C takes place in a concerted step which is very similar to that of route A. In the transition state 7-TS of route C, the bond distances involving the proton transferred to the amine nitrogen from the water oxygen are  $d_{O_{w2}-H} = 1.52 \text{ \AA}$  and  $d_{H-N_{C4'}} = 1.11 \text{ \AA}$  and the bond distances involving the proton transferred from the alcohol to the water molecule are  $d_{O_{C4'}-H} = 1.02 \text{ \AA}$  and  $d_{H-O_{w2}} = 1.64 \text{ \AA}$  (Figure 6B).

The free energies of activation for the water-mediated proton transfer in routes A and C are respectively 10.6 kcal/mol and 15.3 kcal/mol. Such energy difference is mostly due to the higher stabilization of the carbinolamine 6 in route C resulting from the hydrogen bonding between the phosphate group and the water molecule. In addition, the water molecule which participates in the proton transfer exhibits a formal charge of -0.78 (according to a Mulliken's population analysis) in the transition state 7-TS as a result of the asymmetry in the double proton transfer. Therefore, the free energy of activation in route C is higher than that of route A also due to the electrostatic repulsion between the phosphate group and the formal negative charge of the water molecule.

In route B, the proton transfer from the C4' hydroxyl group to the C4' amine takes place in two steps. The first step is similar to the proton transfer of routes A and C. Namely, the C4'-hydroxyl group donates a proton to the water oxygen, which simultaneously donates a second proton to the phenoxide oxygen O3'. In the transition state 7-TS of route B, the distances involving the proton transferred to the water

molecule, are  $d_{\text{O}3'-\text{H}} = 1.27 \text{ \AA}$  and  $d_{\text{H}-\text{O}w2} = 1.52 \text{ \AA}$  while the distances involving the proton transferred to the phenoxide oxygen are  $d_{\text{O}w2-\text{H}} = 1.53 \text{ \AA}$  and  $d_{\text{H}-\text{O}3'} = 1.03 \text{ \AA}$  (Figure 6C). This first step requires an activation free energy of 11.4 kcal/mol, which is lower than that of the proton transfer in route C and similar to that of route A. This step yields the intermediate 8, a different protonation state of carbinolamine 6 which is 8.0 kcal/mol more energetic (Table 2, Figure 3) and exhibits a protonated O3' phenol and an alkoxide at C4' bridged by a water molecule. Additionally, the phenol group in 8 donates a hydrogen bond to the amine nitrogen at C4', arranged for the next step of the reaction.

The proton transfer between the phenol and the amine groups is direct and takes place without participation of any water molecule via a barrierless transition state 9-TS. In this reaction, the O3'-H distance increases from 1.05 Å in 8 to 1.14 Å in 9-TS as the H-N<sub>C4'</sub> distance decreases from 1.54 Å to 1.37 Å (Figure 6D). The calculated activation free energy for this step is slightly negative (-0.7 kcal/mol, Table 2, Figure 3). Since the activation barrier in the potential energy surface at 0 K is only 0.7 kcal/mol, thermal and entropic contributions at 298.15 K are enough to make this reaction proceed without free energy barrier. Furthermore, it should be noted that if nuclear quantum effects were considered for the proton motion, the effective activation energy would be further reduced.

The calculated activation free energies for the different mechanisms of routes A, B and C are in agreement with those calculated for the transamination (40, 41) and Schiff base formation with PLP (42, 43) and PMP (18). The direct proton transfer between the O3' phenol and the amine groups exhibits a free energy barrier between 0.0 and 5.0 kcal/mol (18), while if a water molecule participates in the reaction by bridging the phenol and the amine groups, a slight increase of 0.5-2 kcal/mol (40, 42) in the activation free energy is observed. However, the proton shift between the C4' geminal groups requires a much larger activation energy. An activation barrier of 12.6/14.6 kcal/mol was reported for the water-mediated proton transfer between the C4' amine groups in the transamination reaction in the active site of Ornithine decarboxylase (41). In fact, in the Lys69-PLP Schiff base hydrolysis in the active site of Ornithine decarboxylase, the proton transfer between the C4' hydroxyl and the C4' amine is catalyzed by the thiol group of Cys360 because the water-mediated reaction exhibits a very large barrier of 47 kcal/mol (43).

In these previous reactions, the water molecule that starts the hydrolysis of the Schiff base transfers a proton to the phenoxide group O3' simultaneously to its attack on the imine carbon (18, 42, 43). So, the protonation of the C4' amine exhibits a low activation energy. In the present reaction, the proton is transferred to the alkoxide group at Cp during the water addition to C4'. Therefore, the O3' phenoxide group remains unprotonated and the protonation of the C4' amine requires a larger activation energy because it has to start with the deprotonation of the C4' hydroxyl group, which is much less acidic than the phenol O3'.

In all three reaction routes, the formed compound 10 c corresponds to the zwitterionic counterpart of carbinolamine 6 where the hydroxyl at C4' is deprotonated

and the amine at C4' is protonated. Such zwitterionic state is 5.4, 6.2 and 9.3 kcal/mol more unstable than the neutral species 6 in the routes A, B and C respectively.

### 3.2.4. Schiff base hydrolysis and release of the reduced carbonyl compound

Formation of intermediate 10 in route A weakens the amino bond with the C4' carbon as deduced from the increase in the C4'-N distance between the carbinolamine 6 (1.47 Å) and the intermediate 10 (1.58 Å). Additionally, the unprotonated alkoxide oxygen at C4' in 10 is located in the plane of the pyridine ring (C3-C4-C4'-O dihedral is 175°), so it remains as far as possible from the phenoxide oxygen O3'. As a result, the protonated nitrogen at C4' lies over the pyridine ring (C3-C4-C4'-N dihedral is 61°), forming a hydrogen bond with the oxygen O3' at 2.00 Å. This conformation favors the release of the amino group because it involves the alignment of the C4'-N bond with the p orbitals of the n-system, analogously to the protonation/deprotonation reactions and the addition of carbonyl groups at Ca. In fact, in the transition state of the C4'-N bond cleavage (11-TS), the C3-C4-C4'-N dihedral is 78°. The distance in 11-TS between the C4' carbon and the nitrogen N is 2.25 Å and the forming aldehyde group at C4' is almost coplanar with the pyridine ring (i.e. C3-C4-C4'-O is 172° and C3-C4-C4'-H is 9°). All these conditions contribute to the final hydrolysis step, which exhibits a significantly low activation free energy barrier of 2.4 kcal/mol (Table 2, Figure 3).

The release of the amino acid in route C is very similar to that of route A since the conformations of intermediate 10 are also similar. In the intermediate 10 of route C, the 5'-phosphate group accepts a hydrogen bond from the explicit water molecule which is also hydrogen bonded to the deprotonated oxygen atom of C4'. This oxygen atom lays in the same plane of the pyridine ring (i.e. C3-C4-C4'-O 176°) positioning the protonated amine group over the ring (i.e. C3-C4-C4'-N<sub>im</sub> 61°) at a 1.58 Å of the carbon C4' and forming a hydrogen bond with the oxygen O3' at 1.97 Å. As in route A, this conformation places the leaving group in a most favorable position for its elimination. In fact, the nitrogen atom is at 2.23 Å of C4' carbon and the torsion of C3-C4-C4'-N<sub>im</sub> is 64° in the transition state 11-TS, so the orientation of the amino group barely changes from the reactant 10.

The calculated activation free energy for this last step is -0.1 kcal/mol whereas the potential energy of activation at 0 K including the ZPE corrections is 2.6 kcal/mol. Therefore, the thermal effects and the increase in entropy associated to additional degrees of freedom during the hydrolysis completely eliminate the activation energy at 298 K. However, taking into account the accuracy of DFT calculations and the modeling of solvent effects, it is either possible for this step to involve a positive low activation barrier at 298 K.

Similarly to the activation free energies calculated for routes A and C, the release of the amine compound has been reported to require a very low activation energy for the PLP Schiff bases with methylamine in aqueous solution ( $AG^* = 0.3$  kcal/mol) (42) and the  $\epsilon$ -amino group of Lys69 in the active site of Ornithine decarboxylase ( $AG^* = 1.4$  kcal/mol) (43).

In route B, the intermediate 10 is in a rather different conformation because the alkoxide group of C4' is hydrogen bonded to the oxygen O3' through a bridging water molecule. Accordingly, this alkoxide remains perpendicular to the pyridine ring (i.e. C3-C4-C4'-O4 81°), so the C4'-N bond is not as favorably oriented for its cleavage (i.e. C3-C4-C4'-N 42°) as in routes A and C.

For this step of route B, the activation free energy barrier is significantly larger than that of routes A and C (i.e. 6.8 kcal/mol) but it is still a modest barrier at 298 K for the reaction to take place at a considerable rate. In this case, the higher activation free energy has to be attributed to the higher stability of intermediate 10 with respect to those of routes A and C (Table 2, Scheme 4) due to the hydrogen bond formation involving the alkoxide and phenoxide groups. Additionally, the transition state 11-TS of route C is higher in energy to those of routes A and C because of the electrostatic repulsion between the phenoxide and the forming aldehyde groups which face each other in a *cis*- conformation. Furthermore, such repulsion hinders the forming aldehyde group to accommodate in the plane of the pyridine ring, impeding the delocalization of the p electrons.

In all three routes, the final state 12 is a non-covalent complex formed by reaction products. This state is common to all routes because the conformational barriers that lead to the most stable state are negligible at 298 K. In this complex the released amino acid is placed on top of the pyridine plane at an approximate distance of 3.2 Å measured from the amine nitrogen, Ca carbon and the carboxylate group. Both the amine and carboxylate groups of the amino acid are oriented towards the protonated pyridinium nitrogen in order to stabilize the complex by electrostatic interactions, although the directions in which these groups interact do not permit the formation of hydrogen bonds. However, the 5'-phosphate group does form hydrogen bonds with the explicit water molecule ( $d_{O-H}$  1.71 Å) and with the amine group of the amino acid ( $d_{O-H}$  2.1 Å).

As a first approach, the steady state approximation can be applied to the final intermediate 10 of each route strictly as depicted in Scheme 3. The obtained apparent rate constants for the reaction are  $1.9 \cdot 10^9 \text{ s}^{-1}$ ,  $1.7 \cdot 10^8 \text{ s}^{-1}$  and  $8.2 \cdot 10^9 \text{ s}^{-1}$  for routes A, B and C respectively. In comparison with the rate constants of the reprotonation reactions (Table 1), these results support that carbonyl scavenging reactions occur preferentially, which is in agreement with the experiments of Richard and co-workers (28, 29, 32) and Dalling et al. (33).

#### 4. Implications on Enzyme Catalysis

The crystallographic structures of PLP-dependent enzymes show that the PLP cofactor is surrounded by polar and charged residues in the active site which establish hydrogen bond and salt bridge interactions (26, 27, 44-46). These interactions maintain the Schiff base formed between the PLP cofactor and the substrate amino acid (also known as external aldimine) in the active site, favor the orientation of the appropriate Ca substituent for its heterolytic cleavage, but also play a reactive role in different catalytic stages. In fact, NMR studies corroborate that the enzymatic reactivity of PLP

Schiff bases resembles that of PLP Schiff bases in protic polar solvents (47). Furthermore, a number of particular aspects of PLP-enzymatic catalysis have been understood from the results of experiments and computations of PLP-catalyzed reactions carried out in aqueous solution (18, 28-33, 38, 40-44, 48, 49).

The condensation reaction studied in this work is, in fact, the reverse mechanism of a PLP-catalyzed retro-aldol reaction. Serine hydroxymethyltransferase (SHMT) is an extensively studied PLP-dependent enzyme which catalyzes a reversible retro-aldol/aldol reaction between serine and glycine (45, 50). The accepted mechanism for this reaction involves the participation of tetrahydrofolate as a holder of the formyl group between such amino acids (50). However, SHMT also catalyzes the analogue retro-aldol reactions on other 3-hydroxyamino acids without the participation of tetrahydrofolate (50).

For serine, as well as for other 3-hydroxyamino acids, the cleavage of the Cp-Cot bond in the retro-aldol reaction generates a carbanion at Ca which is stabilized by delocalization of the negative charge across the  $\pi$ -system of the imine and pyridine moieties. Then, this intermediate may evolve by reprotonation at either the Ca or the C4' carbon atoms. Nevertheless, only the protonation at Ca occurs in the SHMT active site.

As reported in previous works, the proton transfer from the phenol oxygen O3' to the imine nitrogen N<sup>im</sup> makes the Ca position 5 pK<sub>a</sub> units more basic than the C4' (30). Additionally, if the phenol and imine groups remain ionized and the pyridine nitrogen is unprotonated, the basicity between Ca and C4' is further increased to 7 pK<sub>a</sub> units (30). Accordingly, and assuming a Brønsted relationship with  $p=1$ , the reprotonation kinetics of the carbanion intermediate is highly favored at Ca over C4' for such protonation states. However, the interactions displayed by the external aldimine in the SHMT active site do not seem to favor the reprotonation of the carbanionic intermediate at Ca with respect to C4'. First of all, the residues interacting with the phenol oxygen O3' are neutral (Ser171 and His199) (Figure 7), which are not as effective as cationic residues to stabilize the phenoxide form of O3', as it is observed in the active site of Alanine racemase (46). Secondly, the Schiff base pyridine nitrogen is presumably protonated since its closest interacting residue is the sidechain carboxyl group of Asp197 (Figure 7). However, it should be noted that the protonated pyridine nitrogen contributes to stabilize the deprotonated phenoxide oxygen O3' and the protonated imine nitrogen N<sup>im</sup> (47, 48). Lastly, the  $\epsilon$ -amino group of Lys226, a conserved residue which selectively reprotonates the carbanionic intermediate either at Ca or C4' respectively in PLP-dependent racemases and transaminases (44, 46), is closer to the C4' carbon than to the Ca one (Figure 5). Therefore, how is the selectivity for the reprotonation at Ca achieved in SHMT?

The data on the pK<sub>a</sub> of PMP- and PLP-Schiff bases (28-31, 48), the protonation barriers calculated in the present work and the crystallographic structures of SHMT complexed with the PLP-Gly Schiff base (45), point out that the sidechain carboxyl group of Glu53 may be the acid catalyst responsible for the reprotonation at Ca. According to the crystal structures, the carboxyl group of Glu53 is in the acidic form (45, 50). In addition, it is properly positioned over the plane of the Schiff base to

transfer a proton to the Ca of the carbanionic intermediate, and it is closer to the Ca (approximately 4.3 Å) than to the C4' carbon (approximately 5.3 Å), which would ensure the selectivity for the former position (Figure 7). Finally, the results reported in this work show that the acid-catalyzed reprotonation of Ca by carboxylic acids is the only reaction faster than the aldol condensation (Figure 2, Figure 4, Table 1, Table 2), which would avoid the reversion of the generated carbanion intermediate back to the 3-hydroxyamino acid form in the active site of SHMT.

## 5. Conclusions

The step involving the aldol condensation, in which the carbonyl species is reduced by formation of a covalent bond with the Ca carbon, exhibits a lower activation free energy barrier than those of the reprotonation reactions of the carbanion by water or most acid catalysts. However, the reprotonation catalyzed by acetic acid is the only faster reaction than the aldol condensation. Therefore, under physiological pH conditions, the formed carbanions of PLP- and PMP-Schiff bases are likely to act as carbonyl scavengers. Furthermore, the reduced carbonyl group is a thermodynamically stable product either as a PLP-Schiff base adduct or after the hydrolysis reaction.

After the aldol condensation, some steps of the hydrolysis require higher activation energies due to the stabilization of the reaction intermediates. However, the free energy of the corresponding transition states is lower than the activation free energy of the condensation step 3-TS (Scheme 3). So, the apparent activation free energy of the overall process corresponds to the activation barrier of the initial step (i.e. ~4 kcal/mol). On the other hand, the activation energies of the reverse reactions for the intermediates are higher than those of the direct reactions, so the formation of the final end-products is also kinetically favored (Figure 4).

The participation of the 5'-phosphate group favors the addition of the carbanion to the carbonyl group in the first step of the reaction and the release of the end-products in the final step of the hydrolysis. Oppositely, the formation of the carbinolamine 6 and its evolution to the zwitterionic state 10 are hindered by the 5'-phosphate group due to the stabilization of intermediates 4 and 6 respectively (Figure 4). Taking into account the high mobility of the 5'-phosphate group in solution at room temperature, it is reasonable that the actual reaction mechanism will not strictly follow one of the proposed routes A, B or C. Instead, the reaction will proceed through the lowest energy path according to the most favorable conformation in each step. Taking everything into consideration, these results support an inhibition mechanism of AGE formation thanks to PLP and PMP Schiffbases which has not been considered so far.

Concerning the reactivity of PLP in enzymes, our results point out that after the cleavage of Ca-Cp bond in the active site of Serine hydroxymethyltransferase, the carbanion intermediate is reprotonated at Ca by a carboxylic acid (i.e. the sidechain of Glu53). In this way, a high selectivity for the reprotonation at Ca in front of C4' is ensured and, simultaneously, the reversion to the formation of the Ca-Cp bond is prevented.



## Supporting Information

Cartesian coordinates, energy and thermodynamic values of the optimized structures.

## Acknowledgments

One of us (R.C.) wishes to acknowledge a fellowship from the Spanish MEC within the FPU program. The authors are grateful to "Centro de Cálculo de Supercomputación de Galicia" (CESGA) and to "Centre de Supercomputació de Catalunya" (CESCA) for access to their computational facilities.

## References

- 1- Maillard, L. C. Action des acides amines sur les sucres: formation des melanoidines par voie methodique C. R. Hebd. Seances Acad. Sci., 1912 154, 66-68.
- 2- Rabbani, M.; Thornalley, P. J. GLycation research in amino acids: a place to call home Amino Acids, 2012 42, 1087-1096.
- 3- Li, Y.; Cohenford, M. A.; Dutta, U.; Dain, J. A. The Structural Modification of DNA Nucleosides by Nonenzymatic Glycation: an in Vitro Study Based on the Reactions of Glyoxal and Methylglyoxal with 2'-Deoxyguanosine Anal. Bioanal. Chem. 2008 390, 679-688.
- 4- Miyazawa, T.; Nakagawa, K.; Shimasaki, S.; Nagai, R. Lipid glycation and protein glycation in diabetes and atherosclerosis Amino Acids, 2012 42, 1163 - 1170.
- 5- Thornalley, P.; Wolff, S.; Crabbe, J.; Stern, A. The autoxidation of glyceraldehyde and other simple monosaccharides under physiological conditions catalyzed by buffer ions Biochim. Biophys. Acta, 1984 797, 276-287.
- 6- Brownlee, M. Advanced protein glycosylation in diabetes and aging Annu Rev Med 1995, 46, 223-234.
- 7- Brownlee, M. Glycation Products and the Pathogenesis of Diabetic Complications Diabetes Care 1992, 15, 1835-1843.
- 8- Baynes, J. W.; Thorpe, S. R. Role of Oxidative Stress in Diabetic Complications. A New Perspective on an Old Paradigm Diabetes 1999, 48, 1-9.
- 9- Bailey, A. J.; Paul, R. G.; Knott, L. Mechanisms of maturation and ageing of collagen Mech. Ageing Dev. 1998, 106, 1-56.
- 10- Kume, S.; Takeya, M.; Mori, T.; Araki, N.; Suzuki, H.; Horiuchi, S.; Kodama, T.; Miyauchi, Y.; Takahashi, K. Immunohistochemical and Ultrastructural Detection of Advanced Glycation End Products in Atherosclerotic Lesions of Human Aorta with a Novel Specific Monoclonal Antibody Am. J. Pathol. 1995, 147, 654-667.
- 11- Grillo, M. A.; Colombatto, S. Advanced glycation end-products (AGEs): involvement in aging and in neurodegenerative diseases Amino Acids 2008, 35, 29-36.
- 12- Ramasamy, R.; Vannucci, S. J.; Shi, D. Y. S.; Herold, K.; Shi, F. Y.; Schmidt, A. M. Advanced glycation end products and RAGE: a common thread in aging, diabetes, neurodegeneration, and inflammation Glicobiology 2005, 15, 16-28.
- 13- Hoyer, S. J. The brain insulin signal transduction system and sporadic (type II) Alzheimer disease: an update J. Neural. Transm. 2002, 109, 341-360.
- 14- Miranda, H. V.; Outeiro, T. F. The sour side of neurodegenerative disorders: the effects of protein glycation J. Pathol. 2010, 211, 13-25.
- 15- Adrover, M.; Vilanova, B.; Muñoz, F.; Donoso J. Inhibition of Glycosylation Processes: the Reaction Between Pyridoxamine and Glucose Chem. Biodiv. 2005, 2, 964-974.

- 16- Adrover, M.; Vilanova, B.; Muñoz, F.; Donoso J. Pyridoxamine, A Scavenger Agent of Carbohydrates Int. J. Chem. Kin. 2007, 3, 154-167.
- 17- Adrover, M.; Vilanova, B.; Frau, J.; Muñoz, F.; Donoso J. A comparative study of the chemical reactivity of pyridoxamine, Ac-Phe-Lys and Ac-Cys with various glycating carbonyl compounds Amino Acids 2009, 36, 437-448.
- 18- Ortega-Castro, J.; Adrover, M.; Frau, J.; Salva, A.; Donoso, J.; Muñoz, F. DFT Studies on Schiff Base Formation of Vitamin B6 Analogues. Reaction between a Pyridoxamine-Analogue and Carbonyl Compounds J. Phys. Chem. A 2010, 114, 4634-4640.
- 19- Voziyan, P. A.; Metz, T. O.; Baynes, J. W.; Hudson, B. G. A Post-Amadori Inhibitor Pyridoxamine Also Inhibits Chemical Modification of Proteins by Scavenging Carbonyl Intermediates of Carbohydrate and Lipid Degradation J. Biol. Chem. 2002, 277, 3397-3403.
- 20- Voziyan, P. A.; Hudson, B. G. Pyridoxamine as a Multifunctional Pharmaceutical: Targeting Pathogenic Glycation and Oxidative Damage Cell. Mol. Life Sci. 2005, 62, 1671-1681.
- 21- Ortega-Castro, J.; Frau, J.; Casasnovas, R.; Fernández, D.; Donoso, J.; Muñoz, F. High- and Low-Spin Fe(III) Complexes of Various AGE Inhibitors J. Phys. Chem. A 2012, 116, 2961-2971.
- 22- Ortega-Castro, J.; Adrover, M.; Frau, J.; Donoso, J.; Muñoz, F. Cu Complexes of Some AGEs Inhibitors Chem. Phys. Lett. 2009, 475, 277-284.
- 23- Adrover, M.; Vilanova, B.; Frau, J.; Muñoz, F.; Donoso, J. The Pyridoxamine action on Amadori compounds: A reexamination of its scavenging capacity and chelating effect Bioorg. Med. Chem. 2008, 16, 5557-5569.
- 24- Casasnovas, R.; Ortega-Castro, J.; Donoso, J.; Frau, J.; Muñoz, F. Theoretical calculations of stability constants and pKa values of metal complexes in solution. Application to Pyridoxamine-Copper(II) complexes and their biological implications in AGE inhibition Phys. Chem. Chem. Phys. Submitted
- 25- Caldés, C.; Vilanova, B.; Adrover, M.; Muñoz, F.; Donoso, J. Understanding non-enzymatic aminophospholipid glycation and its inhibition. Polar head features affect the kinetics of Schiff base formation Bioorg. Med. Chem. 2011, 19, 4536-4543.
- 26- Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes Biochim. Biophys. Acta, 2011, 1814, 1407-1418.
- 27- Eliot, A. C.; Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations Annu. Rev. Biochem. 2004, 73, 383-415.
- 28- Toth, K.; Richard, J. P. Covalent Catalysis by Pyridoxal: Evaluation of the Effect of the Cofactor on the Carbon Acidity of Glycine J. Am. Chem. Soc. 2007, 129, 3013-3021.
- 29- Crujeiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. Substituent Effects on Electrophilic Catalysis by the Carbonyl Group: Anatomy of the Rate Acceleration for PLP-Catalyzed Deprotonation of Glycine J. Am. Chem. Soc. 2011, 133, 3173-3183.

- 30- Casasnovas, R.; Adrover, M.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. C-H Activation in Pyridoxal-5'-phosphate Schiff bases: The Role of the Imine Nitrogen. A Combined Experimental and Computational Study *J. Phys. Chem. B.* 2012, 116, 10665-10675.
- 31- Casasnovas, R.; Frau, J.; Ortega-Castro, J.; Donoso, J.; Muñoz, F. C-H Activation in Pyridoxal-5'-phosphate and Pyridoxamine-5'-phosphate Schiff bases: Effect of Metal Chelation. A Computational Study *J. Phys. Chem. B.* 2013, 117, 2339-2347.
- 32- Toth, K.; Amyes, T. L.; Richard, J. P.; Malthouse, P. G.; NíBeillíú, M. E. Claisen-Type Addition of Glycine to Pyridoxal in Water *J. Am. Chem. Soc.* 2004, 126, 10538-10539.
- 33- Dalling, D. K.; Grant, D. M.; Horton, W. J.; Sagers, R. D. Carbon 13 NMR Study of Nonenzymatic Reactions of Pyridoxal 5'-Phosphate with Selected Amino Acids and of Related Reactions *J. Biol. Chem.*, 1976, 251, 7661-7668.
- 34- Gaussian 09, Revision B.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian, Inc., Wallingford CT, 2009.
- 35- Manerich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions *J. Phys. Chem. B* 2009, 113, 6378-6396.
- 36- Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals *Theor. Chem. Acc.* 2008, 120, 215-241.
- 37- Lide, D. R. Ed. *CRC Handbook of Chemistry and Physics* 87<sup>th</sup> edn. 2007 CRC Press Inc. Boca Raton.
- 38- Griswold, W. R.; Nieto Castro, J.; Fisher, A. J.; Toney, M. D. Ground-State Electronic Destabilization via Hyperconjugation in Aspartate Aminotransferase *J. Am. Chem. Soc.* 2012, 134, 8436-8438.
- 39- Dunathan, H. C. Conformation and Reaction Specificity in Pyridoxal Phosphate Enzymes *Proc. Natl. Acad. Sci. U.S.A.*, 1996, 55, 712-716.
- 40- Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. Density Functional Theory Studies on Transimination of Vitamin B6 Analogues through Geminal Diamine Formation *J. Phys. Chem. A.*, 2004, 108, 11709-11714.
- 41- Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5'-Phosphate-Requiring Enzymes *J. Chem. Theory Comput.*, 2011, 7, 1356-1368.
- 42- Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. DFT Studies on Schiff base Formation of Vitamin B6 Analogues *J. Phys. Chem. A.*, 2003, 107, 9409-9414.

- 43- Oliveira, E. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Mechanism of Formation of the Internal Aldimine in Pyridoxal 5'-Phosphate-Dependent Enzymes *J. Am. Chem. Soc.* 2011, 133, 15496-15505.
- 44- Griswold, W. R.; Fisher, A. J.; Toney, M. D. Crystal structures of Aspartate Aminotransferase Reconstituted with 1-Deazapyridoxal 5'-Phosphate: Internal Aldimine and Stable L-Aspartate External Aldimine *Biochemistry*, 2011, 50, 5918-5924.
- 45- Trivedi, V.; Gupta, A.; Jala, V. R.; Saravanan, P.; Rao, G. S. J.; Rao, N. A.; Savithri, H. S.; Subramanya, H. S. Crystal Structure of Binay and Ternary Complexes of Serine Hydroxymethyltransferase from *Bacillus stearothermophilus* *J. Biol. Chem.* 2002, 227, 17161-17169.
- 46- Watanabe, A.; Yoshimura, T.; Mikami, B.; Hayashi, H.; Kagamiyama, H.; Esaki, N. Reaction mechanism of alanine racemase from *Bacillus stearothermophilus* - X-ray crystallographic studies of the enzyme bound with N-(5'-phosphopyridoxyl)alanine *J. Biol. Chem.* 2002, 277, 19166-19172.
- 47- Sharif, S.; Denisov, G. S.; Toney, M. D. Limbach, H. H. NMR Studies of Coupled Low- and High-Barrier Hydrogen Bonds in Pyridoxal-5'-Phosphate Model Systems in Polar Solution *J. Am. Chem. Soc.* 2007, 129, 6313-6327.
- 48- Casasnovas, R.; Salvà, A.; Frau, J.; Donoso, J.; Muñoz, F. Theoretical study on the distribution of atomic charges in the Schiff bases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms *Chem. Phys.* 2009, 355, 149-156.
- 49- Christen, P.; Metzler, D. E., Eds. In *Transaminases*; John Wiley: New York, 1985.
- 50- Schirch, V.; Szebenyi, D. M. E. Serine Hydroxymethyltransferase revisited *Curr. Op. Biol. Chem.* 2005, 9, 482-487.

Table 1. Calculated activation and reaction free energies at 298.15 K ( $AG^*$ ,  $AG_r$ ) activation and reaction potential energies at 0 K ( $AE^*$ ,  $AE_r$ ) and calculated microscopic kinetic constants ( $k$ ) for the reprotonation reactions of the PLP-Gly Schiff base carbanion by different acid catalysts. Experimental  $pK_a$  values for the acid catalysts. Relevant distances and torsion angles between the atoms involved in the proton transfer reaction measured at the optimized transition state structures.

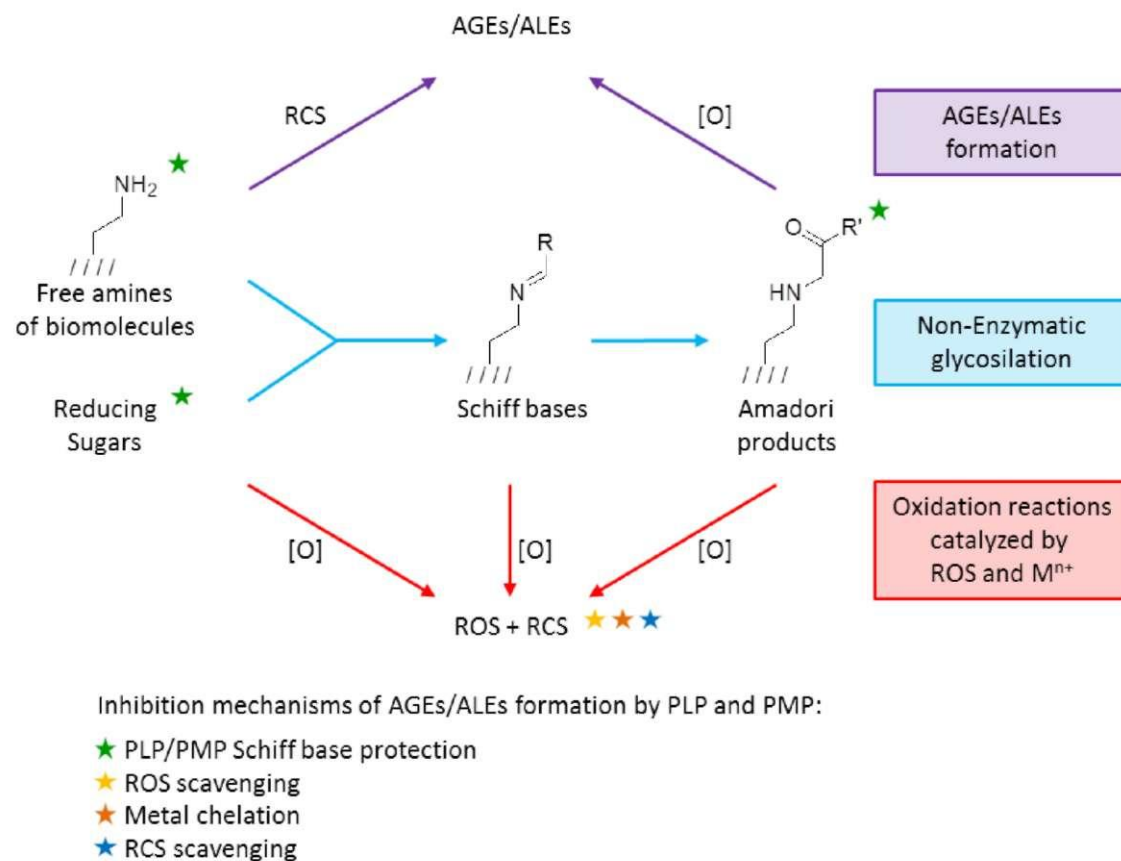
Acid catalyst	extpl.	$AG^*$ ( $AE^{*0}$ ) (kcal/mol)	$AG_r$ ( $AE_r^0$ ) (kcal/mol)	$k$ ( $s^{-1}$ )	dCa-H (Å)	dx-H (Å)	dCa-X (Å)	H-Ca-N <sub>im</sub> -C4' Torsion (deg)
CH <sub>3</sub> COOH	4.76	3.7 (2.6)	-15.4 (-16.2)	$1.2 \cdot 10^{10}$	1.47	1.17	2.64	85.9
H <sub>3</sub> PO <sub>4</sub>	7.21	8.5 (8.1)	-4.3 (-4.4)	$3.6 \cdot 10^8$	1.42	1.21	2.63	90.4
C <sub>6</sub> H <sub>5</sub> OH	9.99	11.0 (10.5)	-4.5 (-3.3)	$5.1 \cdot 10^4$	1.41	1.21	2.59	75.7
CH <sub>3</sub> SH	10.33	9.9 (9.2)	0.2 (0.6)	$3.3 \cdot 10^5$	1.47	1.58	3.02	71.2
CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	10.66	5.9 (5.6)	-8.9 (-8.6)	$2.8 \cdot 10^8$	1.44	1.27	2.71	81.7
CH <sub>3</sub> OH	15.5	15.8 (15.7)	11.9 (12.6)	$1.7 \cdot 10^1$	1.29	1.33	2.61	69.3
H <sub>2</sub> O	15.7	15.7 (15.6)	8.5 (9.6)	$1.8 \cdot 10^1$	1.30	1.34	2.64	78.5

Experimental  $pK_a$  values taken from reference 37. <sup>0</sup>Zero-point Energy Corrections at 0 K included in the potential energies.

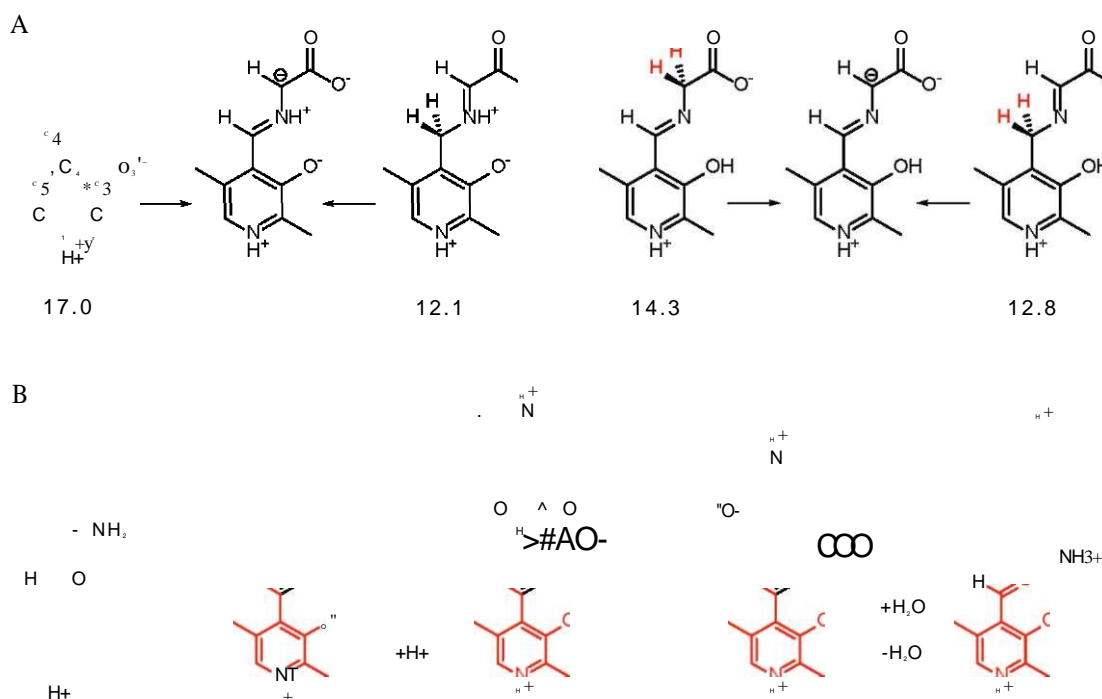
Table 2. Calculated relative free energies at 298.15 K ( $AG$ ) and potential energies at 0 K ( $AE$ ) of each transition state and intermediate structure with respect to the reactants complex 1, and calculated microscopic kinetic constants ( $k$ ) for each reaction step of routes A, B and C.

Structure	$AG_A$ ( $AE_A^0$ ) (kcal/mol)	$AG_B$ ( $AE_B^0$ ) (kcal/mol)	$AG_C$ ( $AE_C^0$ ) (kcal/mol)	$k_A$ ( $s^{-1}$ )	$k_B$ ( $s^{-1}$ )	$k_C$ ( $s^{-1}$ )
1	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)			
2			-2.2 (-2.4)			
3-TS	4.6 (4.4)	4.6 (4.4)	3.9 (2.2)	$2.5 \cdot 10^9$	$2.5 \cdot 10^9$	$8.2 \cdot 10^9$
4	-7.1 (-8.6)	-7.1 (-8.6)	-9.8 (-12.6)			
5-TS	0.6 (-1.5)	0.6 (-1.5)	0.0 (-3.3)	$1.4 \cdot 10^7$	$1.4 \cdot 10^7$	$4.2 \cdot 10^5$
6	-10.7 (-12.4)	-13.7 (-14.4)	-14.2 (-16.4)			
7-TS	-0.1 (-3.3)	-2.3 (-5.4)	1.1 (-2.7)	$9.9 \cdot 10^4$	$3.0 \cdot 10^4$	$3.4 \cdot 10^1$
8		-5.7 (-7.3)				
9-TS		-6.4 (-8.3)			$1.8 \cdot 10^{13}$	
10	-5.3 (-6.9)	-7.5 (-9.3)	-4.9 (-7.9)			
11-TS	-2.9 (-3.9)	-0.7 (-1.6)	-5.0 (-7.1)	$9.6 \cdot 10^{10}$	$6.0 \cdot 10^7$	$8.0 \cdot 10^{12}$
12	-14.8 (-16.3)	-14.8 (-16.3)	-14.8 (-16.3)			

<sup>0</sup>Zero-point Energy Corrections at 0 K included in the potential energies.

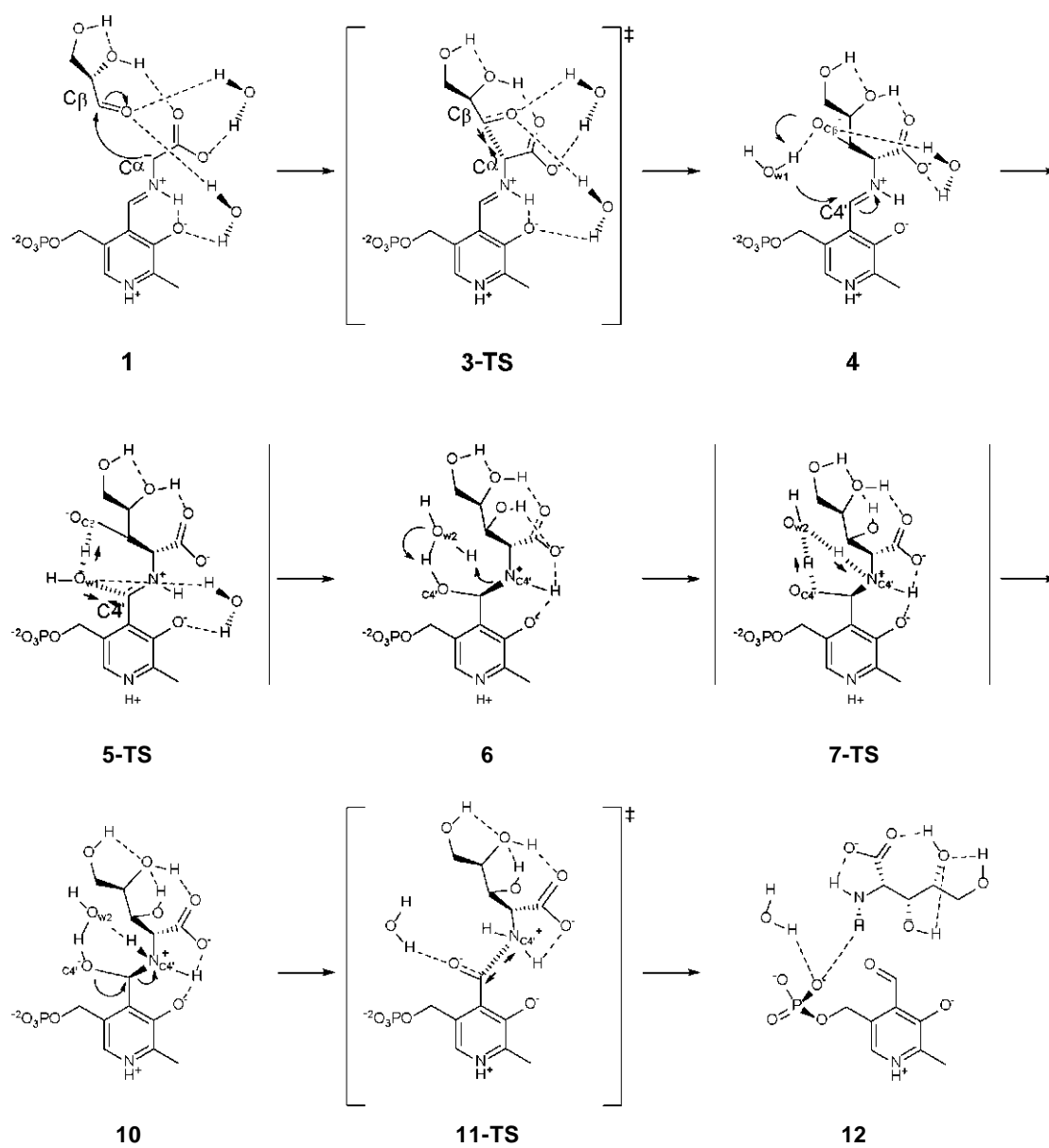


Scheme 1. Formation pathways of advanced glycation and lipoxidation end-products (AGEs/ALEs). The star labels indicate the points of action of PLP and PMP which inhibit the formation of AGEs and ALEs.

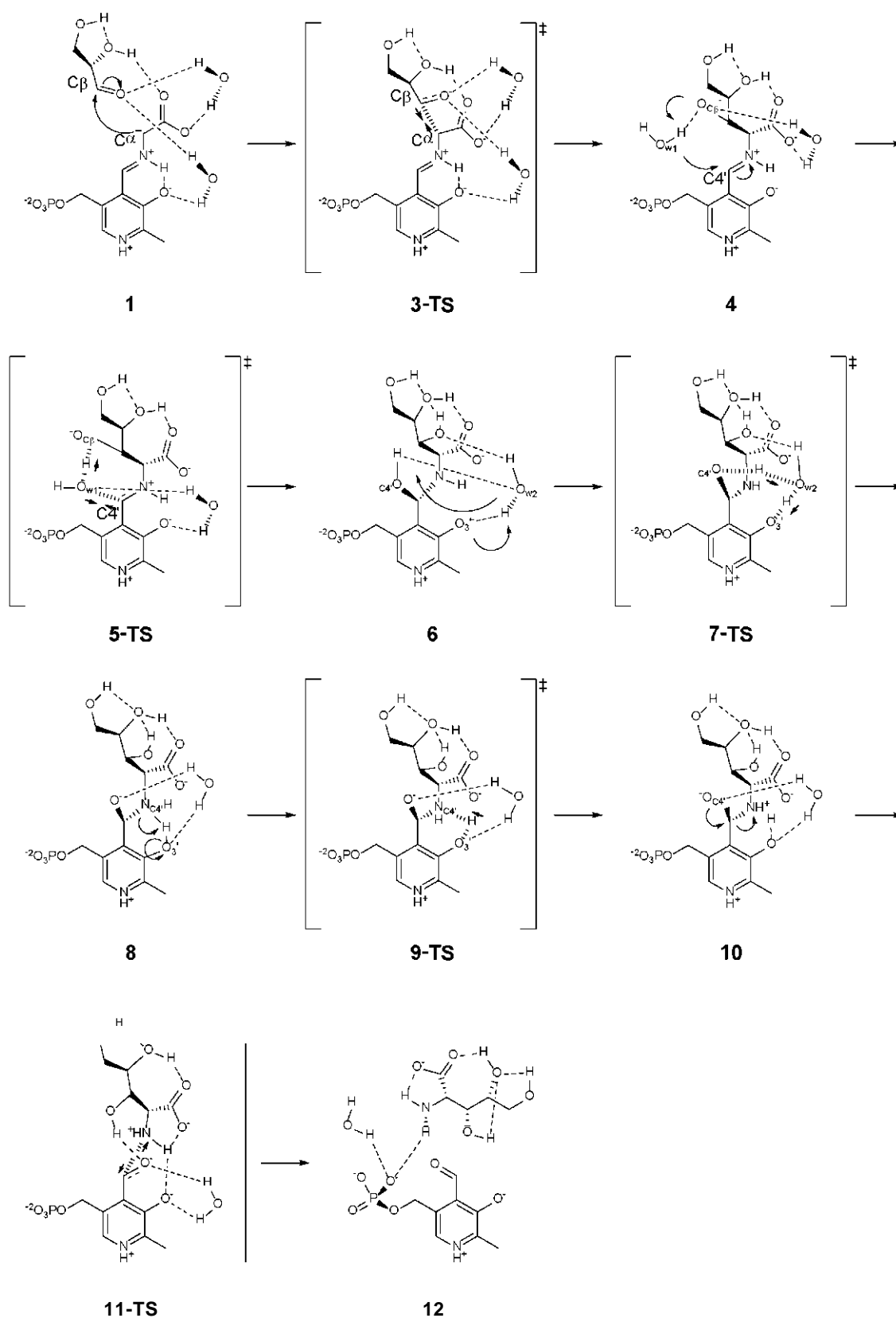


Scheme 2. A)  $pK_a$  values of Ca in the PLP-Glycine Schiff bases and  $pK_a$  values of C4' in the PMP-glyoxylic acid Schiff bases. The depicted values were obtained with B3LYP calculations in reference (30). B) Mechanism of aldol addition of PLP to a PLP-Gly Schiff base reported by Richard and co-workers (28). The reaction starts with the formation of a Schiff base between PLP (red) and glycine, which is later deprotonated. The resulting carbanion reduces the carbonyl group of a second PLP molecule (blue) by acting as a nucleophile. The final hydrolysis releases the catalytic PLP (red) and the glycine linked to the reduced PLP (blue).

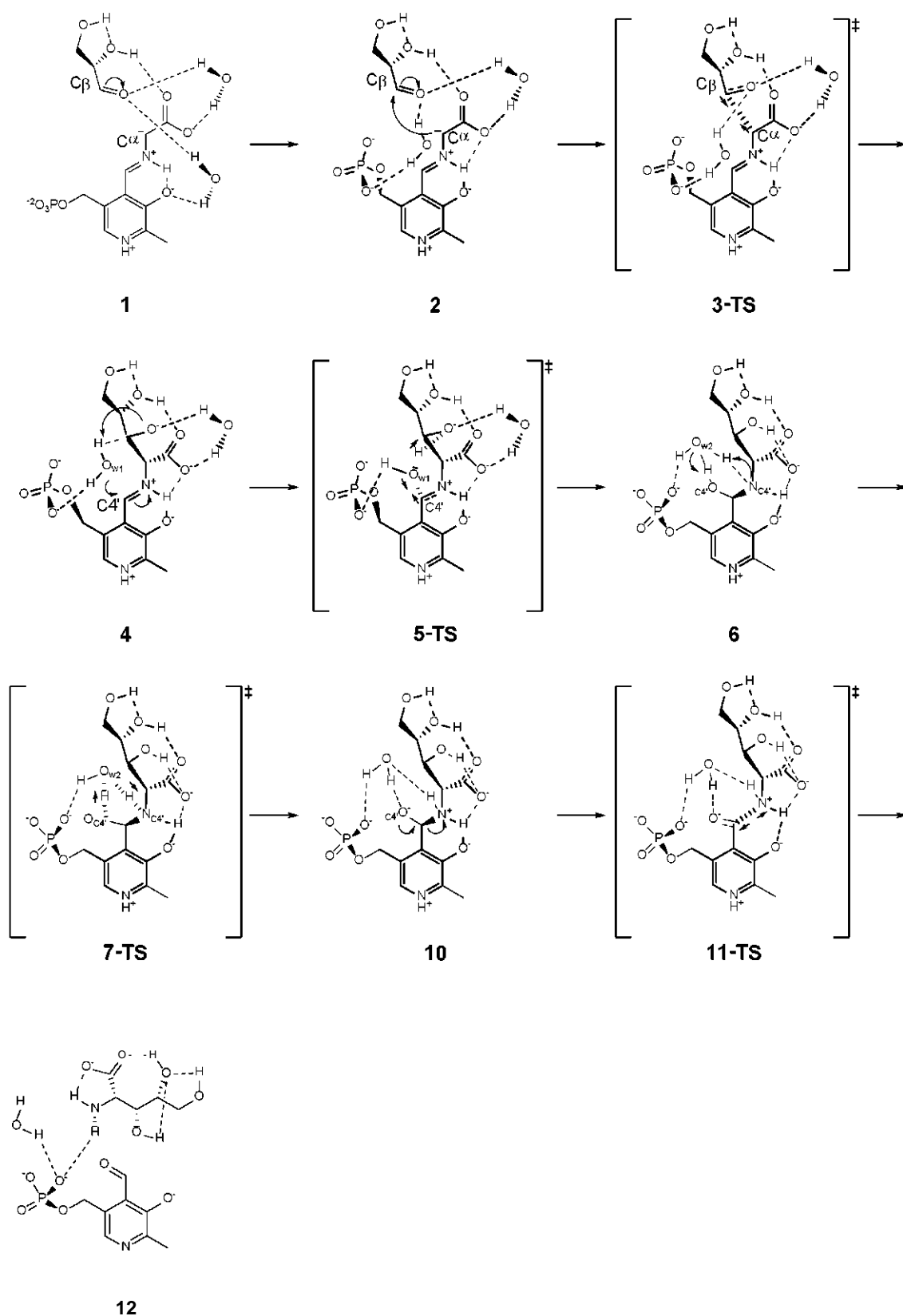




Scheme 3. Reaction mechanism of aldol condensation and hydrolysis of the Schiff base as proposed in route A.



Scheme 4. Reaction mechanism of aldol condensation and hydrolysis of the Schiff base as proposed in route B.



Scheme 5. Reaction mechanism of aldol condensation and hydrolysis of the Schiff base as proposed in route C.

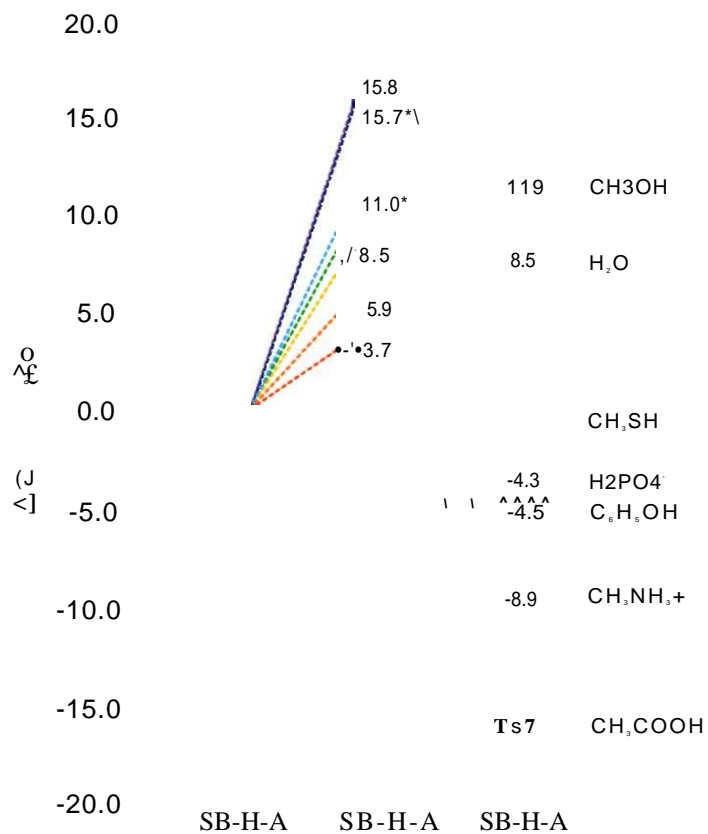


Figure 1. Free energy profiles for the acid-catalyzed reprotonation reactions of PLP-Gly Schiff base carbanions at Ca.

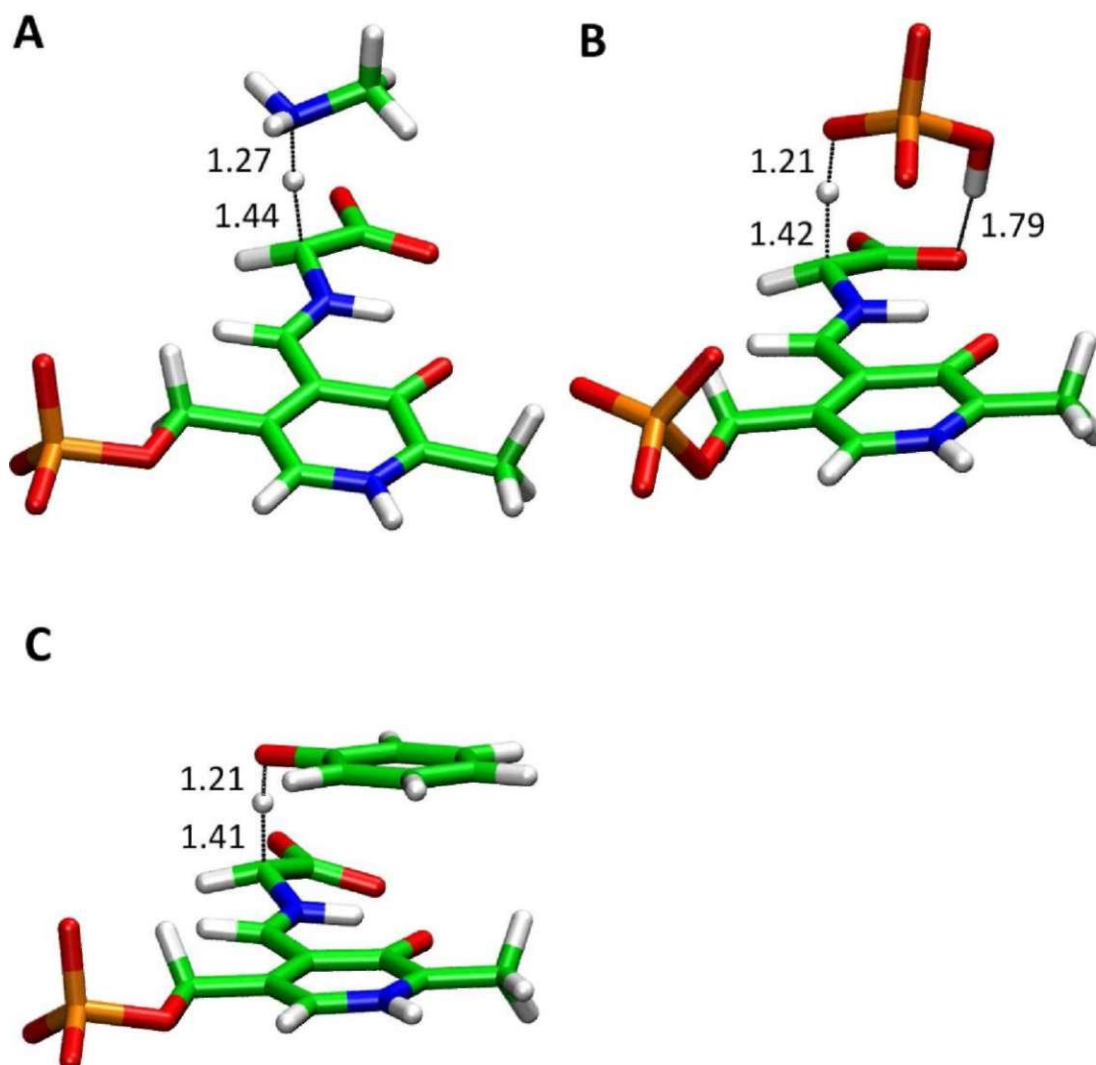


Figure 2. Transition state structures of the proton transfer reactions between the Ca carbon of PLP-Gly Schiff bases and methylammonium (A), dihydrogen phosphate (B) and phenol (C). All distances are measured in Angstrom.

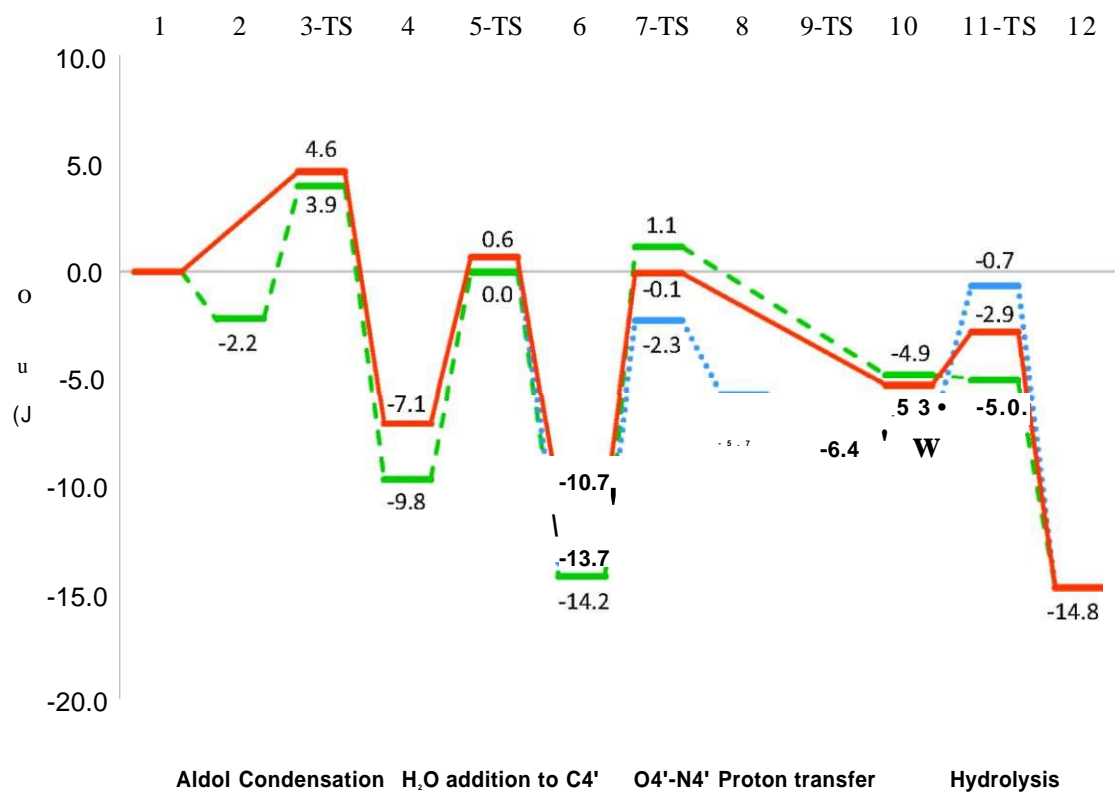


Figure 3. Free energy profiles of the aldol condensation between PLP-Gly carbanion and glyceraldehydes and later hydrolysis of the formed adduct for routes A (orange, solid line), B (blue, dotted line) and C (green, dashed line).

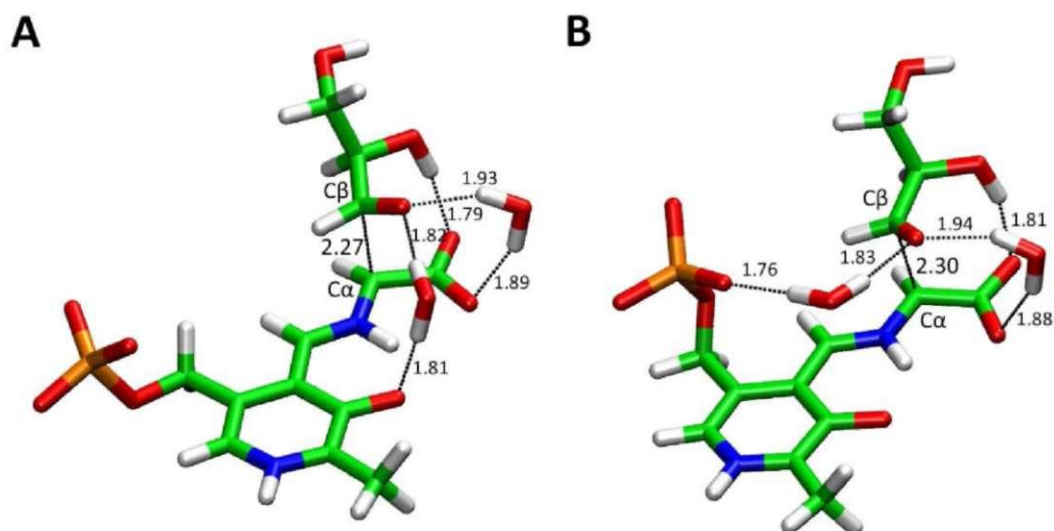


Figure 4. Optimized transition state structures of the aldol condensation between the C $\alpha$  carbon of the PLP-Gly carbanion and the carbonyl carbon (C $\beta$ ) of glycerinaldehyde (3-TS) for routes A and B (A) and C (B). All distances are measured in Angstrom.

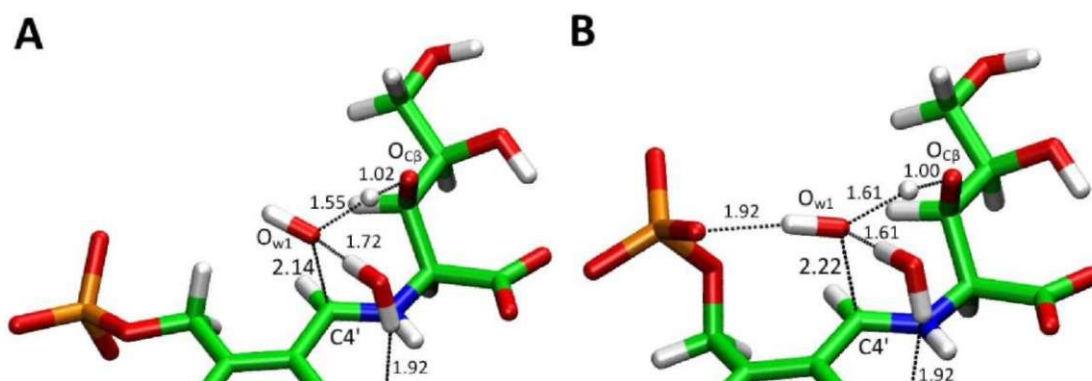


Figure 5. Optimized transition state structures of the water addition to the C $_{4'}$  imine carbon of the aldol condensation adduct Schiff bases (5-TS) for routes A and B (A) and C (B). As observed, the proton transfer from O $_{w1}$  to O $_{C\beta}$  during the first half-reaction generates a hydroxyl anion in situ which is the actual nucleophile species. All distances are measured in Angstrom.

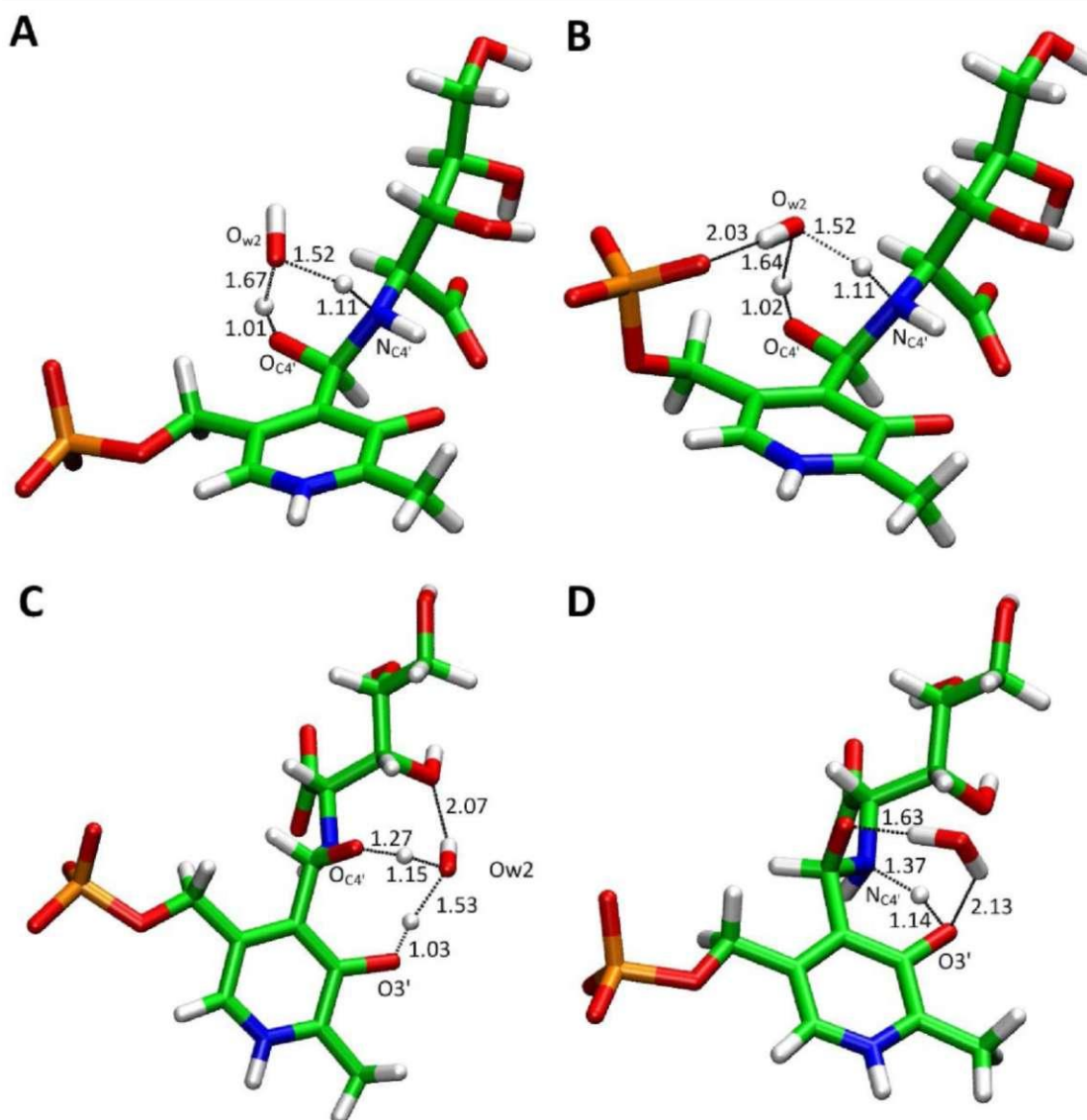


Figure 6. Optimized transition state structures for the proton transfer reaction between the geminal hydroxyl and amino groups of C4'. In routes A (A) and C (B), the proton transfer takes place in a concerted transition state 7-TS with the participation of a reactive water molecule. In (A) and (B), the proton transfer from the water molecule to  $N_{C4'}$  is already completed in the transition state while the proton transfer from  $O_{C4'}$  to the water molecule is barely initiated. In route B the reaction takes place in two consecutive transition states 7-TS (C) and 9-TS (D). The first step, 7-TS, consists in a proton transfer from  $O_{C4'}$  to  $O3'$  with the participation of a water molecule (C). In the second step, 9-TS, the proton is directly transferred from  $O3'$  to  $N_{C4'}$  (D). All distances are measured in Angstrom.



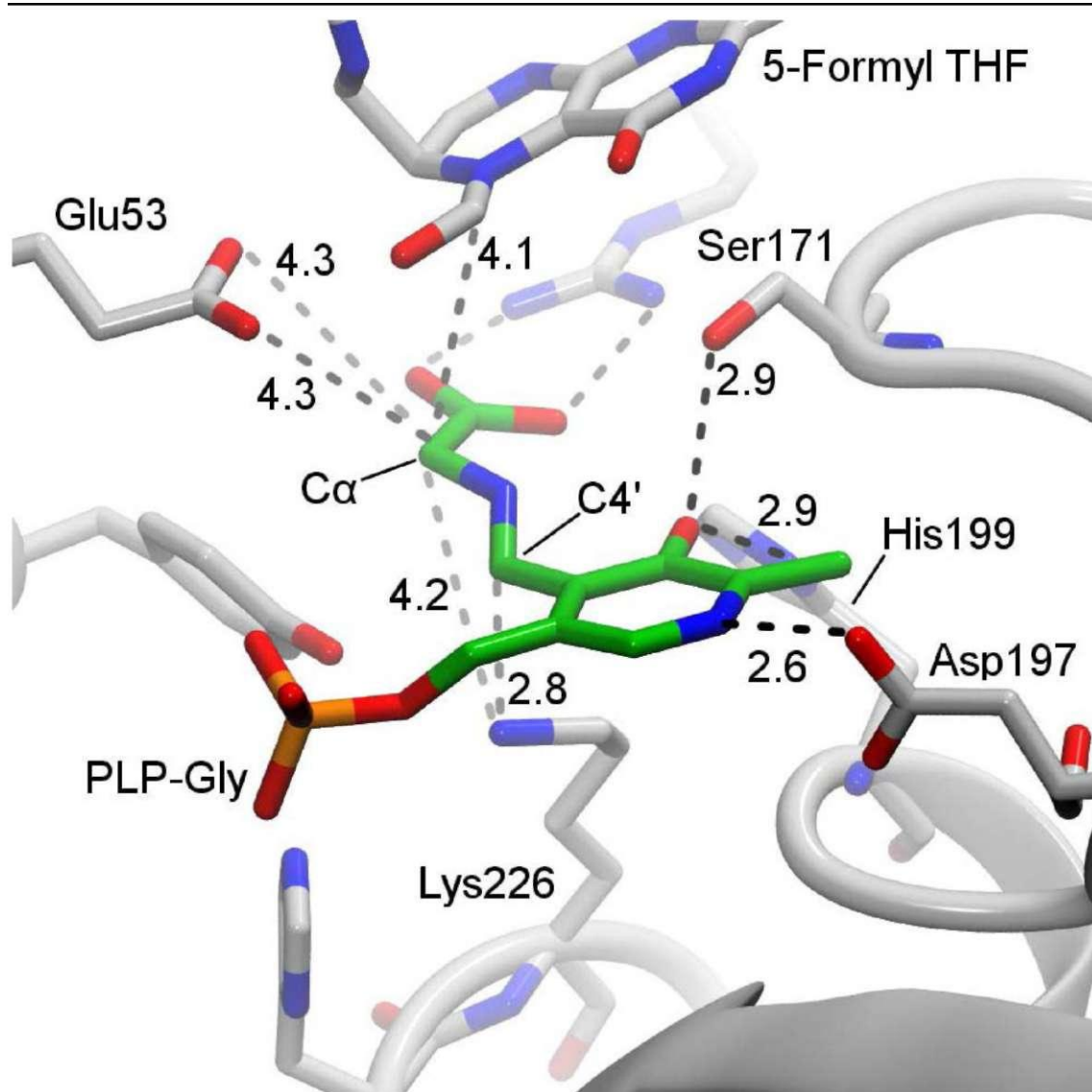
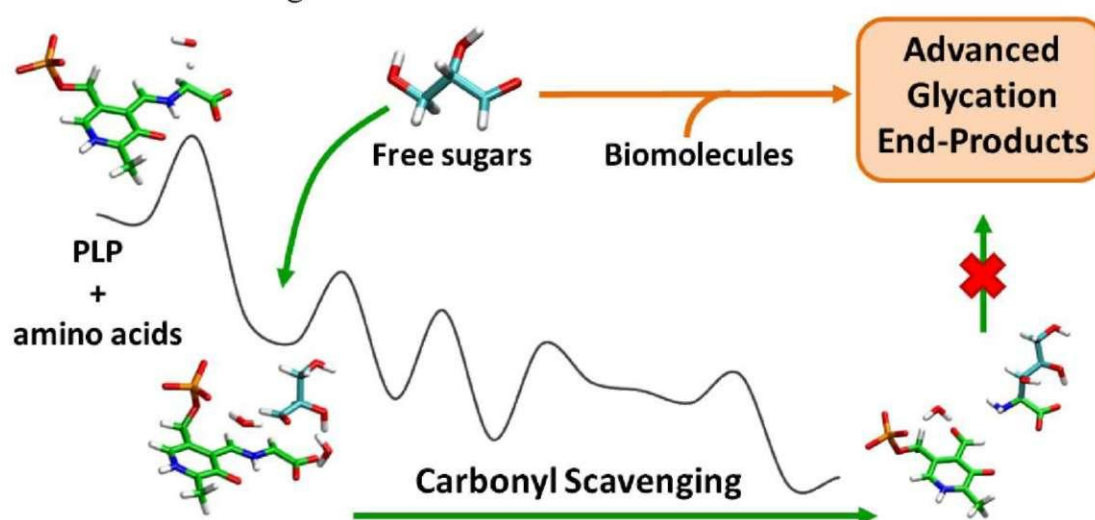


Figure 7. Interactions of the PLP-Gly Schiff base in the active site of Serine hydroxymethyltransferase from *Geobacillus stearothermophilus* (PDB code 1KL2, (45)). All distances are measured in Angstrom.

Table of Contents Image



4.3.2. Extraordinaire decarboxylation rates catalyzed by modestly efficient enzymes. A QM/MM metadynamics study on the enzymatic and nonenzymatic pyridoxal 5'-phosphate-catalyzed decarboxylation of amino acids



**Extraordinaire decarboxylation rates catalyzed by modestly efficient enzymes. A QM/MM metadynamics study on the enzymatic and nonenzymatic pyridoxal 5'-phosphate-catalyzed decarboxylation of amino acids**

Rodrigo Casasnovas, Juan Frau, Sebastiano Caravati, Josefa Donoso, Francisco Muñoz, Michele Parrinello

**Abstract**

Amino acid decarboxylases are among the most efficient enzymes. Almost all of these enzymes require pyridoxal 5'-phosphate (PLP) as cofactor, which forms a Schiff base with the substrate amino acid. We provide insight into the reaction mechanism and catalytic origins of PLP-decarboxylases by studying the PLP-catalyzed decarboxylation of ornithine in the active site of Ornithine decarboxylase (ODC), aqueous solution and gas-phase. The activation and free energies of these reactions were obtained from Density Functional Theory QM/MM metadynamics simulations, supported by QM/QM ONIOM calculations. Our calculations indicate that ODC increases the decarboxylation rate by  $10^9$ -fold. However, the PLP cofactor is responsible for  $10^6$  of the total catalysis, while the exclusive contribution of the enzyme only constitutes the remaining 10-fold acceleration. The active site environment resemblance with the solvent structure after the transition state designates the carboxylate desolvation as an important factor in the enzyme catalysis. In addition, the enzyme structure creates an electrostatic gradient which destabilizes the reactants by placing more anionic than cationic residues closer to the carboxylate group. The enzyme also promotes the protonation of the PLP pyridine nitrogen in the active site, which accounts for  $10^6$  of the cofactor catalysis but it is deprotonated at physiological pH in solution. The gas-phase simulations show that an extra 10-fold acceleration is missed in the active site in favour of maintaining a high energy carbanionic intermediate, in which the proton in the intramolecular hydrogen exhibits high mobility between the imine and phenol groups. By hindering a full protonation of the phenol group, protonation of the intermediate at Ca is favoured in front of transamination side-reactions.

## 1. Introduction

The rates of uncatalyzed biochemical reactions in aqueous solution differ immensely by a factor larger than  $10^{16}$  (1, 2). For instance, the half-time of  $\text{CO}_2$  hydration is approximately 5 s (i.e.  $k \sim 10^{-7} \text{ s}^{-1}$ ) (3), whereas that of phosphodiester anion hydrolysis is approximately 31 million years, which corresponds to a rate constant of  $\sim 10^{-16} \text{ s}^{-1}$  (4). Typically, the bonds that link the monomers of proteins, nucleic acids and polysaccharides exhibit half-times for the spontaneous cleavage ranging  $10^2$ - $10^6$  years. However, such chemical inertness is vital to guarantee the stability of biopolymers under physiological conditions (1).

Amino acids are amongst the most stable biomolecules regarding the reactivity of the bonds formed at the alpha carbon (Ca) and, therefore, spontaneous transformations on this position are extremely infrequent events. Indeed, the reported activation free energies for the proton abstraction of Ca in alanine and for the decarboxylation of glycine are respectively  $\sim 33$  (5) and  $\sim 40$  kcal/mol (6), which correspond to half-times of 5000 years and 1.1 billion years.

In the cell, numerous reactions are linked by sharing common reactants and products in what is known as metabolic pathways. Therefore, in order to be coordinated, all such reactions have to proceed within similar timescales or, have similar reaction rates. Thus, the catalysis of each enzyme should be inversely related to the rate of the uncatalyzed reaction (1, 2).

Since spontaneous amino acid decarboxylation constitutes one of the slowest biochemical reactions, amino acid decarboxylases are amongst the most efficient enzymes (1, 2, 6). It is interesting to note that the vast majority of such decarboxylase enzymes do not achieve their full catalytic power from an exclusively proteic active site, but coenzymes are required instead. Whereas pyruvate is used in *Lactobacillus* Histidine and *E. coli* S-adenosylmethionine decarboxylases, the vast majority of amino acid decarboxylases use a form of vitamin B6, pyridoxal 5'-phosphate (PLP) (7).

In the active site, the PLP coenzyme is covalently bound to the  $\alpha$ -amino group of a lysine residue forming a Schiff base or imine known as internal aldimine (Scheme 1). Prior to the decarboxylation step, the amino acid substrate replaces the PLP linkage with the lysine to form a new Schiff base named external aldimine (Scheme 1). Then, either the Ca-H bond or one of the Ca-C bonds undergoes a heterolytic cleavage yielding a carbanionic compound (Scheme 1).

Several factors contribute to the activation of the Ca and to the evolution of the carbanionic intermediate (8). Dunathan (9) hypothesized that the cleaved bond should be aligned with the  $p$  orbitals of the  $n$  system so that the developing negative charge can be delocalized, lowering the energy of the transition state. Hyperconjugation of the Ca-H bond with the  $n$  system has been also shown to reduce the activation barrier for deprotonation by destabilization of the ground state (10).

In addition to stereoelectronic effects, the protonation state of the heteroatoms in the external aldimine strongly influences the activation of the Ca atom. Schiff base formation with PLP reduces the  $pK_a$  of Ca in the zwitterionic glycine from  $\sim 29$  to  $\sim 23$

in solution (11), and sequential protonation of the heteroatoms also reduces the Ca  $pK_a$  by ~6 units per proton down to  $pK_a$  ~6 (12). With respect to the reaction kinetics, the rate of proton abstraction from Ca in Aspartate aminotransferase decreases dramatically by a factor larger than  $10^9$  when the external aldimine is formed with deaza-PLP (13), i.e. a PLP analogue in which the protonated pyridine nitrogen is substituted by a carbon atom. Apart from Ca-H activation, the protonation state of the carbanionic intermediates also plays a significant role in driving their evolution towards protonation at Ca or C4', that is, controlling the reaction specificity (8, 14, 15).

Concerning the PLP-catalyzed decarboxylation of amino acids, Bach et al. (16) reported from correlated *ab initio* MP2 calculations that the zwitterionic ketoenamine tautomers, in which the pyridine and imine groups are protonated and the phenol and carboxyl groups are deprotonated (Scheme 2), reduce the activation energy by Coulombic destabilization of the ground state. The calculations carried out by Toney (17) point out that the activation enthalpies of decarboxylation of PLP aldimines are lower for the ketoenamine tautomers in aqueous solution as well. Additionally, it was noted that the contribution of the imine group to the resonance stabilization is larger than that of the pyridine ring (17).

In aqueous solution, water solvation of the O3' oxygen stabilizes the phenoxide anion in the PLP aldimines formed with amino acids, making the ketoenamine tautomers predominant regardless of the protonation state of the pyridine nitrogen (18, 19). Oppositely, the neutral enolamines are the most abundant tautomers in organic solvents with low dielectric constants (20). In such environment, protonation of the pyridine nitrogen is required to promote deprotonation of the O3' oxygen since its negative charge is stabilized by delocalization in the pyridine ring. That is, the intramolecular imine-phenol and intermolecular pyridine hydrogen bonds are coupled (20). It was also observed that in low polar conditions, microsolvation by proton donors help stabilize the ketoenamine tautomers (20).

These studies are of great interest to understand the interactions of external aldimines with enzymatic residues, which is the third factor that influences enzyme catalysis and reaction specificity (8). A hydrogen bond interaction between the pyridine nitrogen and a carboxyl group of an acidic residue has been observed in the X-ray structures of many PLP-dependent enzymes (8, 21).<sup>15</sup>N NMR experiments carried out by Toney, Limbach and co-workers (22) on Aspartate aminotransferase showed that the enzymatic environment is similar to polar organic media and that the pyridine nitrogen is protonated. Therefore, the intramolecular hydrogen bond should be zwitterionic in Aspartate aminotransferase as well as in other PLP-dependent enzymes which exhibit the pyridine-carboxyl interaction (22). Conversely, Lin and Gao (23) reported from hybrid QM/MM simulations that the enolimine tautomer is predominant in the active site of L-Dopa decarboxylase. Furthermore, the decarboxylation free energy barrier calculated in the active site is lower for the enolimine tautomers (24).

The interactions between the carboxylate group of the external aldimines and the active site are also important for the catalysis. Phillips and co-workers (25, 26) suggested that in the active site of Ornithine decarboxylase (ODC), the carboxylate

group is buried in a hydrophobic pocket that promotes decarboxylation by favouring the more neutral transition state structure while destabilizing the ionic ground state. In fact, the structure of ODC with the aldimine of the native substrate L-ornithine could have not been crystallized. Instead, only the crystal structure with the aldimine of D-ornithine, in which the carboxylate group is oriented towards the solvent, could be obtained (25). However, there should be another factor for in the enzymatic catalysis besides solvent exclusion of the carboxylate because in Diaminopimelate decarboxylase (DAPDC), whose active site structure is very similar to that of ODC, the carboxylate group of the diaminopimelate aldimine is exposed to the solvent (27).

Ornithine decarboxylase catalyzes the decarboxylation of ornithine to yield putrescine (i.e. 1,4-butanediamine), which is the first step in the biosynthesis of polyamines (28). Polyamines are formed by condensation of two or more diamines, such as putrescine or cadaverine (i.e. 1,5-pentanediamine), and play an essential role in the synthesis of DNA and RNA (28). The inhibition of ODC by  $\alpha$ -difluoromethylornithine has been proved useful in the treatment of cell proliferation-related diseases like the African sleeping sickness (29) or pneumonia in AIDS patients, as well as in combination with other drugs in cancer chemotherapy (30, 31).

In this study, we present the results of Density Functional Theory (DFT) quantum mechanical (QM) and hybrid quantum mechanical/molecular mechanics (QM/MM) simulations on the decarboxylation of the PLP-ornithine aldimine in gas phase, aqueous solution and in the active site of Ornithine decarboxylase (ODC). We have carried out metadynamics simulations to obtain the corresponding decarboxylation free energy profiles to help understand the origins of catalysis in PLP-dependent enzymes. Ab initio metadynamics simulations are a most valuable technique as they also provide an atomic description of the dynamic behaviour of the intramolecular hydrogen bond during the decarboxylation reaction in each environment.

DFT-based ab initio simulations are usually restricted to non-hybrid GGA exchange-correlation functionals because of computational cost reasons. While providing reaction energies with reasonable good precision, such GGA functionals usually underestimate activation barriers. On the other hand, geometry optimization techniques of energy minima and transition states on the potential energy surface can be carried out with more costly and accurate methods at the expense of simplification of the studied models. Therefore, traditional QM and QM/MM (ONIOM) *static calculations* were also carried out to complement the molecular dynamics simulations and to obtain a consistent energetic description of the decarboxylation reaction.

## 2. Computational Details

### 2.1. Preparation of the model systems

The X-ray structure of human Ornithine decarboxylase (PDB code 2O00) crystallized at 1.90Å resolution (32) was chosen to study the enzymatic decarboxylation. Ornithine decarboxylase is a homodimer with two identical active sites



located at the interface formed by the N-terminal domain of one dimer and the C-terminal domain of the other (Figure 1). The N-terminal domain adopts a p/a-barrel structure and the C-terminal domain a P-sheet structure, and the binding site of the PLP coenzyme is mostly located by residues of the N-terminal domain (25, 26, 32).

The SWISS-MODEL Workspace server (33) was used to build the missing sections of the protein. The crystal structure of 2OO0 contains the free PLP coenzyme and the 3-aminoxy-1-propane inhibitor in the active site besides crystallized acetate and cadaverine (i.e. 1,5-diaminopentane) out of the active site. Such compounds were removed and the PLP-Ornithine aldimine was modelled by superimposing the active site coordinates of 2OO0 with those of 1F3T Ornithine decarboxylase crystallized with the PLP-Putrescine aldimine (34), which corresponds to the products state of the studied decarboxylation reaction. Then, the carboxylate group was added to the *re* face of the Ca carbon, in the hydrophobic pocket formed by the aromatic sidechains of Phe397 and Tyr389 and the methylene groups of the sidechains of Cys360 and Lys69. In the 2OO0 structure, the thiol group of Cys360 is oriented towards the PLP coenzyme as in the products state of the decarboxylation reaction (25). Therefore it was rotated out of the active site, in the proper conformation of the reactants state (25).

The AmberTools1.2 package was employed in the preparation of the enzyme and solution systems for the molecular dynamics simulations. All histidine residues were protonated on the S2 nitrogen after inspection of the local hydrogen bond patterns with the neighbour residues. A total of 28 Na<sup>+</sup> ions were added to neutralize the protein-ligand, and the entire system was solvated with SPC-FW waters (35) in a box whose edges were at least 7Å away from the protein (~24250 water molecules). For the study of decarboxylation in aqueous solution, the PLP-Ornithine was solvated with ~1180 SPC-FW water molecules in a cubic box. The Amber99SB force field (36) was used in the description of the protein and ions. It is worth to note that the TIP3P force field (37) is the one recommended for water in combination with Amber force fields. However a massive colored-noise generalized Langevin equation thermostat (38) was used for the QM/MM simulations and, therefore, any of the degrees of freedom of any atom in the simulation could be constrained.

The mechanical parameters for the PLP-Ornithine aldimine were obtained from the General Amber Force Field (39) by using the Antechamber module in AmberTools1.2 (40). In order to obtain the atomic charges, the geometry of the PLP-Ornithine aldimine was optimized at the M06-2X/6-31+G(d,p) level in water modelled by the continuum solvent SMD (41) with Gaussian09 (42). Then, the HF/6-31G(d) wavefunction and electrostatic potential were calculated on the optimized geometry to fit the atomic charges according to the Merz-Singh-Kollman scheme (43).

## 2.2. Classical Molecular Dynamics simulations

Previous to the QM/MM molecular dynamics simulations, the enzyme and aqueous solution systems were equilibrated to 300K and 1.0bar by performing classical molecular dynamics simulations with the Sander module in the Amber10 package (44).

Periodic boundary conditions were used during the energy minimizations and molecular dynamics simulations. A cutoff of 10 Å was applied to the calculation of the nonbonded interactions and the Particle Mesh Ewald method (45) was used to calculate the long-range interactions.

The water molecules in the enzyme system were minimized for 3000 steps by using the steepest descent algorithm followed by 7000 additional steps by using the conjugate gradients algorithm. A harmonic restraint with  $k=100$  kcal/molÅ<sup>2</sup> was applied on the protein and ligands atoms to eliminate the solvent-protein clashes. Then, the constraints were removed and the whole system was minimized for another 3000 plus 7000 steps of steepest descent and conjugate gradients algorithms respectively.

The temperature was equilibrated to 300 K by performing 50 ps of Langevin dynamics with a collision frequency of 5.0 ps<sup>-1</sup> in the NVT ensemble. Harmonic restraints with  $k=20$  kcal/molÅ<sup>2</sup> were imposed to the protein and ligands atoms during the heating. The SHAKE algorithm was used to constrain the hydrogen distances so that the time step could be set to 2 fs. Therefore, the water molecules were not flexible at this stage.

The pressure was equilibrated to 1.0 bar in the NPT ensemble by using a Berendsen barostat (46) with a relaxation time of 2 ps. The NPT equilibration was carried out in successive steps in which the restraints on the protein and ligand atoms were gradually reduced to relax the initial crystal structure in solution. Specifically, 100 ps were run by applying harmonic restraints with  $k=20$  kcal/molÅ<sup>2</sup>, 200 ps were run with  $k=0.5$  kcal/molÅ<sup>2</sup>, 100 ps with  $k=0.05$  kcal/molÅ<sup>2</sup> and the last 100 ps were run with  $k=0.001$  kcal/molÅ<sup>2</sup>. Finally, all restraints on the protein and ligands were removed, together with all the constraints in the hydrogen-involving distances and the entire system was equilibrated for 2.5 ns with a time step of 0.5 fs in the NPT ensemble to an average density of 1.06 g/cm<sup>3</sup>. The measured root mean square deviation (RMSD) of the backbone atoms during the NPT stage without constraints resulted 1.4±0.2 Å with respect to the crystallographic structure. In addition, the RMSD of the residues within 10 Å of the external aldimine, including the sidechain atoms, was 0.83±0.06 Å. The RMSD values remained stable during the last ~1.0 ns of the equilibration (Figure S1, Supporting Information), indicating the structural equilibration of the protein. In both cases, the obtained RMSD values fall within the resolution of the crystallographic experiments (i.e. 1.90 Å) (32).

In the case of the PLP-Ornithine aldimine in aqueous solution, the temperature and pressure were equilibrated to 300 K and 1.0 bar for 1 ns of Langevin dynamics with a collision frequency of 1.0 ps<sup>-1</sup> in the NPT ensemble with a Berendsen thermostat (46) with a relaxation time of 1.0 ps. As for the enzyme system, periodic boundary conditions and a cutoff of 10 Å together with the Particle Mesh Ewald method for the long-range part of the nonbonded interactions were used. A time step of 0.5 fs was used and no constraints were imposed on the hydrogen distances to simulate the flexible behaviour of PSC-FW waters. The resulting average density in the NPT equilibration was 1.00 g/cm<sup>3</sup>. However, since the QM/MM simulations will be performed in the NVT ensemble, the simulation box was resized to the average volume obtained during the

NPT equilibration and an additional NVT simulation was performed for 1 ns to relax the solvent.

### 2.3.QM and QM/MM Molecular Dynamics simulations

The CP2K package (47) was used in the QM simulations carried out in gas phase as well as in the QM/MM simulations of the enzyme and solution systems. All the atoms of the PLP-Ornithine aldimine were taken into account in the gas phase model and were all treated at the ab initio level according to the Density Functional Theory (DFT) in the Quickstep module (48, 49) of CP2K. The BLYP exchange correlation functional (50, 51) was used in combination with a double zeta basis set with polarization functions (DZVP) and an auxiliary set of planewaves with a density cutoff of 300 Ry and the core electrons were described with the norm-conserving Goedecker-Teter-Hutter (52) pseudopotentials. A threshold of  $1 \cdot 10^{-6}$  Ha for the energies was used as a criterion of wavefunction convergence in each step of the simulations.

To simulate the gas phase environment, the Blochl method (53) was used to decouple the electrostatic interactions with the periodic images. Since a condition of Blochl's decoupling scheme is that the periodic images only interact electrostatically, the convergence of the interaction energy with respect to the box size was evaluated. Eventually, a box larger than  $16.0 \times 16.0 \times 16.0 \text{ \AA}^3$  was required for a correct decoupling of the periodic replicas.

During the decarboxylation reaction, all the atoms involved in the cleavage of the Ca-COO bond are part of the PLP-Ornithine aldimine, while the water molecules in aqueous solution and the residues in the enzyme active site establish non-bonding interactions with the carboxylate group. The solvent molecules and protein residues also interact with the rest of the aldimine acid/base groups, stabilizing the protonation state of the aldimine so, the protonation state of the aldimine is presumed to be stable along the reaction. Accordingly, in aqueous solution and in the enzyme systems, the QM region was restricted to the PLP hydroxypyridine ring and the reactive ornithine atoms (Scheme 3). On the other hand, since the phosphate and amino groups of the PLP-Ornithine aldimine are simply electrostatically involved in the decarboxylation reaction, they were considered as part of the MM region (Scheme 3). Therefore, in the QM/MM calculations, the PLP C5' and the ornithine Cy were substituted by hydrogen link atoms according to the IMMOM scheme (54), and a scaling factor was used to project the forces of the capping hydrogens on the frontier link MM atoms.

The QM/MM simulations in CP2K make use of the Quickstep and Fist modules for the calculation of the QM and MM regions, while the QM/MM interaction energy and forces are calculated with the QM/MM module according to the formulation of Laino et al. (55, 56).

The protein and counterions were modelled with the Amber99SB force field and the water molecules with the SPC-FW force field as in the classical equilibrations. The smooth Particle Mesh Ewald method (57) and a cutoff of 10Å for the calculation of the real part of the nonbonded interactions were used. All the parameters controlling the calculation of the wavefunction and electron density of the QM region are equal to those

described for the gas phase system with the exception of the box size. The Blochl's decoupling scheme was also used to decouple the periodic replicas, but a smaller minimum size box of  $14.0 \times 14.0 \times 14.0 \text{ \AA}^3$  was required for a correct decoupling because the QM regions considered in solution and in the enzyme are smaller than that considered in gas phase.

An integration time step of 0.5 fs was used in all the QM and QM/MM simulations, which were performed in the NVT ensemble at a temperature of 300 K controlled by a colored-noise generalized Langevin equation thermostat (38). The thermostat parameters were obtained from an online repository (GLE4MD, 58) and were chosen to minimize the correlation time of the potential energy of the vibrational modes with frequencies in the range  $0.4\text{--}4000 \text{ cm}^{-1}$ , which involve the typical frequencies of proteins, water and organic molecules. The PLP-Ornithine aldimine was equilibrated in gas phase, in the ODC active site and in aqueous solution for 3.0 ps, 6.0 ps and 10.0 ps respectively until the average temperatures of the QM and MM systems were stabilized at 300 K.

#### 2.4. Metadynamics

The well-tempered (59) version of metadynamics (60), together with the use of multiple walkers (61) was used with CP2K to obtain the free energy profiles of the decarboxylation reactions. The Ca-COO distance was chosen as the collective variable whose associated potential energy is modified along the metadynamics simulation by the addition of gaussian potentials, commonly known as *hills*. Such hills were deposited every 60 steps (i.e. every 30 fs) with a width of 0.10 Bohr and an initial height of 0.0020 Hartree. An external potential acting on the collective variable was positioned at 7.5 Bohr with a quadratic constant of  $15.7 \text{ kcal/mol\AA}^2$  to avoid an infinite elongation of the Ca-COO distance during the metadynamics simulations. The fictitious temperature at which the collective variable was sampled in the well-tempered metadynamics was set to 4500 K.

Four walkers were used in the gas phase and aqueous solution metadynamics simulations, while six walkers were used for the enzyme reaction. The initial structures of each walker were randomly selected from their respective QM or QM/MM equilibration stages. Since the calculations of each walker run on independent machines and only the information of the hills is shared, the final simulation times of each walker are different. Nevertheless, in the case of the gas-phase reaction the metadynamics simulation was performed at least for 42 ps for each walker, while in the case of the aqueous solution and enzyme reactions, each walker was simulated at least for 62 ps and 30 ps respectively. Eventually, the total simulation time was 182 ps, 256 ps and 212 ps for the gas phase, aqueous solution and enzyme metadynamics.

In the transition state, the probability of evolving to reactants and products is equal. Therefore, if a large enough number of molecular dynamics simulations is initiated with random initial velocities, the more the initial structure resembles the transition state, the closer the distribution of trajectories reaching reactants and products would be to 50:50. This procedure, also known as committor analysis, allows the verification that the

maximum of the free energy profile corresponds to the transition state of the studied reaction. We carried out the committor analysis for the reactions in solution, enzyme and gas-phase, and the accepted structures resulted in a reactant/product distribution of 49:51, 50:50 and 45:55, respectively, over a total number of 100 trajectories in each case (Figure S2).

## 2.5. Geometry optimization of critical points on the potential energy surface

Geometry optimizations were carried out to find the energy minima (i.e. reactants and products states) and the transition state structures of the decarboxylation reactions in the gas phase, aqueous solution and in the active site of Ornithine decarboxylase.

The PLP-Ornithine models for the reactions in aqueous solution and in the ODC active site were obtained from the equilibrated systems by classical molecular dynamics simulations. In the case of the aqueous solution reaction, the entire PLP-Ornithine aldimine plus ten water molecules, which correspond to the first solvation shell of the pyridine, phenol and carboxylate groups, were taken into account. In addition to the explicit water molecules, the SMD continuum solvent model (41) was used for the calculations of the reaction in aqueous solution to model the bulk effects of water beyond the first solvation shell.

For the enzyme reaction, an active site model of 297 atoms was built by taking into account the entire PLP-Ornithine aldimine and all the residues and water molecules with which establishes direct interactions (Figure 1). Specifically, the model contains the atoms in the residues Arg277-Glu274 from the first monomer and in the residues Cys360, Asp361, Phe397, Asn398 from the second monomer and the sidechains of Lys69, Asp88, Arg154, His197, Ser200, Asp332, Tyr331 and Tyr389 from the first monomer. Additionally, a total of 15 water molecules were included, most of which are found in the crystal structure of ODC, PDB entry 2OO0: two water molecules interacting with the guanidinium group of Arg154, four waters are shared between Arg154 and Glu274, another molecule that only interacts with Glu274, two more waters interacting simultaneously with Lys69 and Asp88, two water molecules interacting with the phosphate group of PLP, another molecule that interacts with the 5-amino group of the ornithine sidechain and three water molecules interacting with Asp361 (Figure 1). The frontier carbons resulting from the elimination of the rest of the protein were saturated with hydrogen atoms.

All the calculations were carried out with the Gaussian 09 package (Gaussian 09). The geometry optimizations were carried out by using the BLYP (50, 51) and M06-2X (62) exchange correlation functionals with the 6-31+G(d) basis set. Vibrational analyses were carried out on the optimized structures to characterize them as energy minima or transition states respectively by the absence of imaginary frequencies or by the presence of a single imaginary frequency, which corresponds to the vibrational motion associated to the decarboxylation reaction. In a post-optimization step, the energies of the optimized structures were refined increasing the basis set to 6-311++G(d,p) with the BLYP and M06-2X functionals. The free energies were calculated as the summation of

the thermal corrections calculated in the vibrational analyses with the 6-31+G(d) basis set and the potential energies calculated with the 6-311++G(d,p) basis set.

An ONIOM scheme (63, 64) was used for the study in the active site model. DensityFunctional Theory calculations with the BLYP and M06-2X functionals were used for the PLP-Ornithine aldimine, while the AM1 (65) semiempirical method was used for the protein and water molecules. All the alpha carbons of the complete residues and the frontier carbons in the fragmented residues were frozen to maintain the active site architecture. The geometry optimizations were carried out in two steps. First, the water molecules were optimized with loose convergence criteria. In the second step, the water molecules and the previously mentioned carbon atoms were frozen and the aldimine and rest of the active site atoms were fully optimized. During the calculation of the refined energies of the active site, the SMD solvent model was used to approximate the long range effects of the missing protein residues and solvent molecules.

### 3. Results and Discussion

#### 3.1. Catalysis of decarboxylation in ODC

The activation free energy of decarboxylation for the PLP-Ornithine aldimine in solution is approximately 17.5 kcal/mol according to the metadynamics simulations (Table 1, Figure 2). A very similar value is obtained via static geometry optimization calculations performed with the same exchange-correlation functional, BLYP (Table 1). As mentioned earlier, non-hybrid GGA functionals tend to underestimate activation energies of chemical reactions. In fact, equivalent calculations performed by using the M06-2X functional point out that the activation free energy in solution is 22.6 kcal/mol (Table 1). The uncatalyzed decarboxylation of glycine in solution requires an activation energy of ~40 kcal/mol (6) (Table 1). Considering that the decarboxylation activation energy of ornithine in solution is similar, the formation of the Schiff base with PLP reduces the barrier by 17.4-22.5 kcal/mol, which involves a  $10^{13}$ - $10^{16}$  fold acceleration of the reaction rate.

The transition state structures obtained from the committor analysis point out that the carboxylate group is eliminated perpendicularly to the plane formed by the pyridine and imine moieties (Figure 3). According to the analysis of electron density, this conformation allows the electron pair generated in the cleavage of the C-COO bond to be stabilized by delocalization across the  $\pi$  system (Figure 3).

The experiments performed by Zabinski and Toney (66) indicate that the decarboxylation activation energy of  $\alpha$ -aminoisobutyrate catalyzed by PLP in solution at pH 5 is 24.8 kcal/mol (Table 1), higher to that calculated in this work for the PLP-Orn aldimine. Recent  $^{15}\text{N}$  NMR experiments indicate that the pyridine nitrogen of PLP Schiff bases with amino acids is deprotonated above pH 4 (19). Since the protonation of this nitrogen atom favors the delocalization and stabilization of the negative charge generated in the transition state, the activation energy of decarboxylation at pH 5 is

expected to be higher than the one calculated for the pyridine nitrogen-protonated aldimine in this work.

It has been reported that the decarboxylation rate of 2-methyl-2-aminomalonate aldimine with PLP shows a 20-fold decrease upon pH increase from 5 to pH 8 (66). Considering that a similar decrease holds for the PLP- $\alpha$ -aminoisobutyrate aldimine, its decarboxylation free energy of activation at pH 8 would be  $\sim 26.6$  kcal/mol (Table 1). The interest of this value is that the protonation state at pH 8 and the one considered in this work only differ in the pyridine nitrogen. Consequently, the contribution to the decarboxylation catalysis by protonation of this nitrogen atom can be estimated from the difference between the activation energy at pH 8 and the one calculated in this work. Consequently, the protonation of the pyridine nitrogen reduces the activation energy by 9.1 kcal/mol and 4.0 kcal/mol respectively, according to the metadynamics simulations and the M06-2X calculations. Since these values only account for 40% or 20% of the total energy barrier reduction according to the metadynamics and M06-2X calculations, the protonation of the pyridine nitrogen is a relevant contribution but not the major one to the catalysis by PLP. Similar conclusions were reported from previous calculations carried out to study the protonation effect of PLP Schiff bases on the decarboxylation barriers (17).

Instead of the major protonation state observed at physiological pH, the decarboxylation study in solution was carried out in this work by considering the protonation state corresponding to the ODC active site. This allows the total enzyme catalysis to be divided in a contribution resulting from Schiff base formation with the cofactor and another ascribed to the proteic environment. According to the metadynamics simulations, the decarboxylation in the ODC active site exhibits 13.8 kcal/mol activation free energy (Table 1, Figure 2), which is in agreement with kinetic experiments carried out by Brooks and Phillips (67) who reported a value of 14.5 kcal/mol. The static calculations carried out with the BLYP and M06-2X functionals in the active site model provide activation energies that follow the same trends observed for the decarboxylation in solution (Table 1). Again, considering the activation energy of the uncatalyzed decarboxylation in solution (i.e.  $\sim 40$  kcal/mol (6), Table 1), ODC causes a remarkable  $10^{19}$ -fold increase in the decarboxylation rate or, alternatively, a reduction of 26.2 kcal/mol in the activation energy or, alternatively.

As in solution, the carboxylate group in the ODC active site is eliminated perpendicularly to the plane ring so that the developing negative charge generated in the transition state is delocalized across the  $\pi$  system (Figure 3). These results point out that the decarboxylation mechanism is mainly governed by stereoelectronic factors inherent to the PLP Schiff base rather than by the environment. Independently of the calculation method, the reduction in the activation energy due to the transfer of the PLP-Orn aldimine from solution to the ODC active site is approximately 3-4 kcal/mol.

Although the carboxylate group of D-Ornithine is solvent-exposed once the external aldimine is formed in ODC, its decarboxylation has been experimentally observed (25, 27). Indeed, the measured decarboxylation rate of the D isomer is approximately  $10^4$ -fold slower than that of the L isomer (27), which corresponds to a  $\sim 4.5$  kcal/mol higher activation energy, in excellent agreement with our results. Similarly, the enzyme

diaminopimelate decarboxylase carries out the elimination of a solvent-oriented carboxylate group on its natural substrate (27). Our results show that the catalysis in these cases is still possible because the determining factor is the Schiff base formation with the pyridine-protonated PLP cofactor. Lin and Gao (24) have also reported analogue conclusions from QM/MM simulations for the PLP-catalyzed decarboxylation of L-Dopa in the enzyme L-Dopa decarboxylase and in solution. At this point, it should be reminded that the PLP-Schiff bases in solution at physiological pH are less protonated than in the ODC active site. So, altogether, the catalysis by ODC involves a  $\sim 10^9$ -fold acceleration with respect to the nonenzymatic reaction catalyzed by PLP in solution, including both the contribution of the enzymatic environment and the pyridine nitrogen protonation in the active site.

Based on kinetic studies of mutant ODCs and crystal structures of ODC complexed with diverse ligands, Phillips and co-workers (25, 26, 68) proposed that the decarboxylation is favored in the ODC active site as a result of desolvation of the carboxylate group. In agreement with this hypothesis, our simulations in ODC show that the carboxylate group of PLP-Orn is positioned in a hydrophobic pocket formed by the aromatic rings of Phe397, Tyr389 and the methylene groups of Lys69 sidechain. Additionally, the carboxylate group does not interact with any water molecules and the only hydrogen bond interaction is formed with the sidechain amide NH<sub>2</sub> group of Asn398 at an average distance of 3.33 Å (Table 2). An equivalent amide group that is present in the PLP-dependent Dialkylglycine decarboxylase active site has been reported to help the carboxylate group adopt the proper conformation for its elimination from Ca (8).

In order to study both the intra- and intermolecular variations upon decarboxylation, the structures from the trajectories of the metadynamics simulations have been clustered in two categories that correspond to the reactants and products regions. The used criterion consists on evaluating whether the Ca-COO distance is shorter or longer than that obtained from the committor analysis for the reactions. Likewise, the radial distribution functions of the PLP-Orn aldimine with the water oxygen (O<sub>w</sub>) and water hydrogen (H<sub>w</sub>) atoms in solution have been obtained separately for the reactants and products regions (Figure 4, Figure S3).

In the reactants state, the first solvation shell of each carboxylate oxygen consists of 3 water molecules in average. The water oxygens (O<sub>w</sub>) are localized in a region between 2.42 and 3.26 Å of the carboxylate oxygens and the most probable O-O<sub>w</sub> distance is 2.77 Å (Figure 4). These water molecules that constitute the first solvation shell situate an average of 2.9 hydrogen atoms (H<sub>w</sub>) at a hydrogen bond distance (i.e. 1.86 Å) of each carboxylate oxygen (Figure 4). In order to quantify the number of hydrogen bonds, two geometrical variables were evaluated. Firstly, the water oxygen atom O<sub>w</sub> should be at 3.5 Å or less of the carboxylate oxygen and, secondly, the angle formed by the O<sub>w</sub>-H<sub>w</sub> and O<sub>w</sub>-O bonds should be equal or less than 30°. According to these criteria, the number of hydrogen bonds oscillates between 1 and 3, reaching a maximum of 4 for some structures, and resulting an average of 2.2 (Table 2).

In the products region, the negative charge is delocalized along the  $\pi$  system of the Schiff base and the eliminated CO<sub>2</sub> is neutral and nonpolar. A consequent decrease to



an average of 0.15 is observed for the hydrogen bonds formed in the first solvation shell between the solvent waters and each CO<sub>2</sub> oxygen. The oxygen atoms O<sub>w</sub> of these water molecules in the first solvation shell occupy a significant larger region (i.e. 2.56-4.73 Å) (Figure 4). In addition the most probable O-O<sub>w</sub> distance increases to 3.26 Å, which is shorter than the most probable O-H<sub>w</sub> distance (i.e. 3.89 Å, Figure 4), indicating that the H<sub>w</sub> hydrogens reorient towards the bulk solvent and do not interact with the CO<sub>2</sub> molecule.

The radial distribution function of the carboxylate and water oxygens in the ODC active site indicates that the carboxylate group is not stabilized by hydrogen bonding interactions (Figure 4). Indeed, comparing the solution and active site radial distribution functions, it is observed that the active site environment is closer to that of the free CO<sub>2</sub> in solution (Figure 4). These results support the hypothesis that the decarboxylation is promoted in the ODC active site by destabilization of the reactants state in favor of the more nonpolar transition state.

It is also worth noting that at physiological pH, ODC exhibits more negatively charged than positively charged residues, resulting in a total negative charge of 28 au. Furthermore, the ODC architecture places the negatively-charged residues closer to the carboxylate group of the PLP-Orn aldimine than the positively-charged ones (Figure 4), creating an electrostatic gradient that destabilizes the reactant state.

The activation barrier in gas-phase corresponds to the intrinsic reactivity of the PLP Schiff bases formed with amino acids and allows a quantification of the maximum catalysis by desolvation. In gas phase, both the transition state and the resulting products of decarboxylation are strongly stabilized with respect to the reactants state due to their lower polarity (Figure 2, Table 1). According to the metadynamics simulations and the static calculations, the activation barrier is 2.4 kcal/mol and 1.4 kcal/mol respectively (Table 1), which is significantly lower than that of the decarboxylation in aqueous solution and also in ODC (Figure 2 Table 1). It is remarkable that the reduction of the activation barrier in ODC is only ~3 kcal/mol of a potential ~15 kcal/mol decrease because of desolvation.

Besides desolvation, other factors, which are difficult to quantify but worth mentioning, contribute to the catalysis in ODC. The most stable conformation of the PLP-Orn aldimine in the ODC active site exhibits the carboxylate group almost perpendicular to the pyridine ring (i.e. the COO-Ca-Nim-C4' torsion equals 98°, Table 2). This conformation involves the alignment of the COO-Ca bond with the orbitals of the  $\pi$  system. Contrarily, in aqueous solution and gas-phase, the most stable conformation places the carboxylate group slightly off the Schiff base plane (i.e. the COO-Ca-Nim-C4' torsions equal -138° and -123° respectively Table 2). In this conformation, the closest carboxylate oxygen to the imine hydrogen is at 2.49 Å in solution and at 2.55 Å in gas-phase so that the iminium and carboxylate group stabilize each other by Coulomb interactions. During the metadynamics simulations, the energy addition to the COO-Ca bond causes the COO-Ca-Nim-C4' dihedral to rotate to average values of 101° in the ODC active site, -110° in solution and -86° in gas-phase prior to decarboxylation (Table 2). Therefore, part of the catalysis in ODC results from

the destabilization of the reactants state in the active site, which forces a more favorable conformation to the transition state. Previous calculations performed by Toney (17) estimated a ~13 kcal/mol conformational destabilization in gas-phase for the aldimines in the same protonation state than the one considered in the present study.

The destabilization of the reactants in the ODC active site is also manifested in the equilibrium Ca-COO bond distance. In solution, where the negatively charged carboxylate group is stabilized by solvation, the average bond length is 1.58 Å (Table 2), whereas the equilibrium length increases to 1.69 Å in gas-phase (Table 2). The average bond length in ODC is 1.65 Å (Table 2), closer to the gas-phase and longer than in solution. As shown by Griswold et al. (10), the Ca-H bond order decreases by a 20% when oriented perpendicularly to the PLP ring due to hyperconjugation with the  $\pi$  system. Similarly, the labilization of the Ca-COO bond in ODC may be caused by similar stereoelectronic effects.

According to the committor analysis, the Ca-COO bond lengths in the transition state in solution, ODC and gas-phase are respectively 2.38 Å, 2.81 Å and 2.13 Å, which match the maxima of the free energy profiles obtained from the metadynamics simulations (Figure 2). These values fulfill the Hammond postulate since the decarboxylation is less endergonic in solution than in ODC, while in gas-phase the reaction is exergonic (Figure 2).

### 3.2. Destabilization of the carbanionic intermediate and reaction specificity of its evolution

It is worth to note that whereas the transition state is stabilized with respect to the reactants state in ODC, the carbanionic intermediate is destabilized (Figure 2). Similar destabilization has also been reported for the carbanionic intermediate resulting from Ca deprotonation in Alanine racemase (69). The absence of protonation at the pyridine nitrogen in this enzyme hinders the delocalization of the negative charge from Ca across the  $\pi$  system. Consequently, the increase in energy of the intermediate favors its fast reprotonation at Ca and, simultaneously, transamination side reactions are kinetically avoided (69). Our results suggest that ODC may take advantage of the same strategy to drive the protonation of the intermediate to Ca. In fact, kinetic experiments of ODC show that the reprotonation of the intermediate is the fastest step in the reaction (67), which supports the presence of a high energy carbanionic intermediate. However, contrarily to Alanine racemase, the interaction between the Schiff base with Glu274 (25, 26) and the detection of a quinonoid-type carbanion intermediate (67) suggest that the Schiff base is protonated at the pyridine nitrogen in ODC. Therefore, two questions arise; namely, 1) What causes the destabilization of the intermediate in ODC? and 2) Which mechanism ensures the reprotonation of the intermediate at Ca?

In the metadynamics simulations, the Glu274 residue belongs to the classical region and no proton transfer is therefore possible from the Schiff base pyridine nitrogen. Consequently, the average  $N_{\text{pyr}}\text{-H}$  and  $\text{H-OGlu274}$  bond lengths show values of 1.04 Å and 1.70 Å that remain invariant during the reaction (Table 2). However, this is not only

a result of the artificial QM/MM division because proton transfers are neither observed in the ONIOM active site models, which are treated completely quantum mechanically. Therefore, the destabilization of the carbanionic intermediate is not likely to be caused by deprotonation of the pyridine nitrogen in the active site during the reaction.

Neither the 5'-phosphate of the PLP nor the 8-amino of the ornithine sidechain interact with the carboxylate or any other groups of the aldimine during the decarboxylation. In solution, these groups remain away of the reaction region with solvation spheres that are also unaltered during the decarboxylation (Figure S3). In gas-phase, the 5-amino group of the ornithine sidechain donates a proton to the 5'-phosphate group during the first picosecond of the equilibrium simulation. Afterwards, the ornithine sidechain extends so that the 5-amino group remains away from both the 5'-phosphate and the carboxylate groups (Figure S3, Table S1). Similarly, the 5'-phosphate and the 5-amino groups are anchored during the entire reaction in the ODC active site. The 5'-phosphate accepts hydrogen bonds from two water molecules, the sidechains of His197B, Ser200B, Tyr389B and with the backbone NH groups of Gly237B, Gly276B and Arg277B (Table S1). These results point out that neither the 5'-phosphate nor the 5-amino group is responsible for the destabilization of the carbanionic intermediate in ODC.

During the simulations in solution, as well as in ODC and gas-phase, the Schiff base adopts a planar conformation that allows the formation on an intramolecular hydrogen bond with the phenol oxygen, which is in agreement with recent NMR studies (19). As a result of this intramolecular interaction, the imine group is not solvated by water molecules in solution and does not establish hydrogen bonding interactions with any residue in the active site.

In solution, the first solvation shell of the phenol oxygen O3' is formed by ~2.3 water molecules at an average distance of 2.77 Å. The number of hydrogen bonds accepted by the O3' oxygen oscillates between 1 and 2, and the average is 1.1 (Table 2). Once the decarboxylation takes place and the generated negative charge delocalizes across the  $\pi$  system to the pyridine ring, the negative charge of the O3' oxygen is also increased. In response, the number of water molecules in the first solvation shell rises to ~2.6. More interestingly, the hydrogen atoms of these water molecules reorient towards O3', the average O3'-H<sub>w</sub> distance decreases from 1.86 Å to 1.79 Å and the average number of hydrogen bonds increases to 1.8 (Table 2).

The interactions at O3' have strong influence on the tautomerism between the phenol and imine groups (18, 20). Our simulations show that in solution and in gas-phase, respectively 97% and 99% of the structures exhibit a protonated imine nitrogen and a deprotonated phenol oxygen. That is, the major tautomer in the reactants state corresponds to the ketoenamine form (Scheme 2). As mentioned previously, the negative charge of O3' increases by delocalization after the transition state. The result in gas-phase is that the proton is mostly transferred from the imine group to O3' since only 11% of the structures show a protonated imine nitrogen (Figure 5). Oppositely, in solution, 75% of the carbanionic intermediate structures remain protonated at the imine

nitrogen because the reorganization of solvent molecules stabilizes the increase of negative charge at O3' and hinders the proton transfer (Table 2, Figure 5).

Similarly to solution and gas-phase, in the ODC active site, 96% of the reactants structures are protonated at the imine nitrogen. After decarboxylation, 40% of the carbanionic intermediate structures are protonated at the imine nitrogen (Figure 5). Therefore, the proton transfer to O3' increases with respect to solution but without reaching the extent observed in gas-phase (Figure 5). In the ODC active site, the O3' interacts with a single water molecule at 2.92 Å, also observed in crystal structures of ODC (25, 26), which establishes an average of 0.8 hydrogen bonds in the reactants and products (Table 2). These results indicate that the carbanionic intermediate in ODC is not stabilized either by a full protonation of O3', as in gas-phase, or by an increase of the hydrogen bonds interactions, as in solution.

The simulations carried out by Lin and Gao (24) on the decarboxylation of PLP Schiff bases show a lower decarboxylation barrier for the O3'-protonated tautomers. Reasonably, part of the catalysis in PLP-dependent decarboxylases could be attributed to a coupling between the decarboxylation reaction and the imine-phenol proton transfer (24). The present metadynamics simulations support this hypothesis since such coupling is clearly observed in gas-phase and, in a lesser extent, in the ODC active site, whereas almost no coupling occurs in solution (Figure 5).

Recent NMR studies indicate that the imine-phenol proton transfer exhibits a high-barrier transition state in the PLP Schiff bases (18, 20). Our data point out that the mobility of the proton in the intramolecular hydrogen bond increases once the carbanionic intermediate is formed (Figure 5), which implies a decrease in the barrier of the proton transfer. In addition, the simulations suggest that the imine-phenol tautomerism of the carbanionic intermediates is easier to modify by specific interactions at O3' than in the Schiff bases. Even though more quantitative studies should be carried out on the tautomerism of the carbanionic intermediates, the results in the present work support previous studies that propose that the interactions of O3' are modulated in the active site of each PLP-dependent enzyme to promote either the enolimine (i.e. O3'-protonated) tautomer or the ketoenamine (i.e. N<sub>im</sub>-protonated) tautomer, which ultimately control the relative rates of protonation of the intermediate at Ca or at C4' (8, 11, 13-15, 69-71). Overall, the degree of protonation and the hydrogen bonds interactions observed for O3' in the ODC active site suggest a compromise between the reduction of the transition state decarboxylation to favor the catalysis and the destabilization of the carbanionic intermediate to control the specificity of the protonation at Ca.

Nevertheless, even the partial protonation of O3' increases the probability of transamination side-reactions. Therefore, the role of the acid catalyst residue is also important to achieve the correct reaction specificity. Once the PLP-Orn aldimine is formed, the *s*-amino group of Lys69 moves away from the Schiff base carboxylate group and forms hydrogen bonding interactions with Asp88. However, in this position, it is still closer to C4' than to Ca (Table 2) so, if Lys69 were the acid catalyst in the protonation of the carbanionic intermediate, the entropic factor would favor the

protonation at C4' over Ca. In fact, several crystal structures of ODC and kinetic experiments of mutant enzymes indicate that Cys360 is the actual acid catalyst in the protonation of the carbanionic intermediate (25, 26, 34). During the metadynamics, the elimination of the carboxylate group as CO<sub>2</sub> involves the loss of the hydrogen bond with the amide-NH<sub>2</sub> group of Asn398 (Table 2). As a consequence, the mobility of the amide group of Asn398 increases and its hydrogen bond with the thiol group of Cys360 is also broken (Table 2). Therefore, the decarboxylation reaction indirectly liberates the thiol group of Cys360 allowing its later rotation inwards the active site, close to Ca for its protonation. Contrarily, the interactions between Lys69 and Asp88 remain unaltered upon decarboxylation, indicating that it is probable that the release of the Lys69 sidechain cannot occur until the protonation of Ca by Cys360. In such mechanism, the deprotonated Cys360 thiol group would be regenerated by neutralization with the Lys69 s-amino group, activating this latter residue as nucleophile for the hydrolysis and release of the final products.

As depicted in Figure 2, the free energy of the transition state in ODC is reduced almost to equalize the free energy of the carbanionic intermediate. Another reason to avoid greater desolvation is that a more hydrophobic active site would probably hinder the Schiff base formation with ornithine and the hydrolysis of the final products. These reactions involve the nucleophilic attack of the incoming amino group of the positively charged iminium group of the already formed PLP Schiff base, as well as a number of proton transfer reactions between ionic/polar groups (72-75). In fact, kinetic studies of ODC reveal that no greater desolvation is needed because the rate-limiting step is the hydrolysis of the final products instead of the decarboxylation step (26, 34, 67, 68).

#### 4. Conclusions

Similarly to other PLP-dependent decarboxylases with their natural substrates, Ornithine decarboxylase accelerates the decarboxylation of L-Ornithine by 10<sup>19</sup>-fold compared to the uncatalyzed reaction in solution, which corresponds to one of the largest catalytic efficiencies for biological reactions. The enzymatic environment reduces the direct interactions with the carboxylate group almost to the maximum and reduces the activation barrier by destabilization of the reactants with respect to the transition state. At physiological pH, ODC bears more negatively charged than positively charged residues. The negatively charged residues are also closer to the aldimine carboxylate, creating an electrostatic gradient which also destabilizes the reactants and favors the elimination of CO<sub>2</sub>. In addition, the conformation of the external aldimine in the ODC active site is ideal from the electronic standpoint for the stabilization of the negative charge generated by the Ca-COO bond cleavage. Eventually, the proton transfer from the imine nitrogen to the phenol oxygen O3' during or after the decarboxylation is more favorable than in solution, which helps stabilize the generated negative charge in the transition state. However, the catalytic contribution exclusively due to ODC is only 3-4 kcal/mol, or a reaction rate acceleration of 10<sup>2</sup>-10<sup>3</sup>-fold, which involves that the remaining 10<sup>16</sup>-fold acceleration originates in the Schiff

base formation with the PLP cofactor. Nevertheless, the ODC active site makes possible the protonation of the PLP pyridine nitrogen with respect to solution, which constitutes a  $10^7$  factor of the total  $10^{16}$ -fold acceleration by the cofactor.

There are two reasons that explain the fact that only a small fraction of the potential catalysis due to desolvation is recovered in the ODC active site. Firstly, the decrease in the activation barrier is enough to make the hydrolysis of the products, and not the decarboxylation, the rate-limiting step in the global kinetics. Furthermore, the nucleophilic attack of Lys69 and the proton transfer reactions that constitute the hydrolysis would be further hindered in a more hydrophobic environment. Secondly, the proton of the intramolecular hydrogen bond in the carbanionic intermediates of PLP aldimines shows great mobility in ODC. Larger desolvation of the aldimine in the active site would involve a large stabilization of the carbanionic intermediate by a higher degree of protonation of O3', which ultimately would favor transamination-side reactions. That is, the reaction specificity is prioritized in front of the decarboxylation catalysis. Finally, in agreement with previous hypothesis, our results indicate that the CO<sub>2</sub> elimination liberates the Cys360 thiol group to act as the acid catalyst in the protonation of the carbanionic intermediate exclusively at Ca while retaining the  $\sigma$ -amino group of Lys69.

## 5. Acknowledgements

The authors acknowledge Michele Ceriotti and Teodoro Laino for helpful comments on methodological and technical aspects of the simulations. The authors are grateful to "Centro de Cálculo de Supercomputación de Galicia" (CESGA) and to "Centre de Supercomputació de Catalunya" (CESCA) for providing access to their computational facilities. R.C. wishes to acknowledge a fellowship from the Spanish MEC within the FPU program.

## 6. Supporting Information Material

Root Mean Square Deviations of the protein during the equilibration stage. Evolution of the Ca-COO bond length of the transition states structures obtained from the committor analyses in solution, ODC active site and gas-phase. Radial distribution functions of the heteroatoms in the PLP-Orn aldimines in aqueous solution. Evolution of the Ca-COO, O3'-H and Nim-H bond lengths during the metadynamics simulations. Average bond lengths, interaction distances and standard deviations in solution, ODC active site and gas-phase.

## 7. References

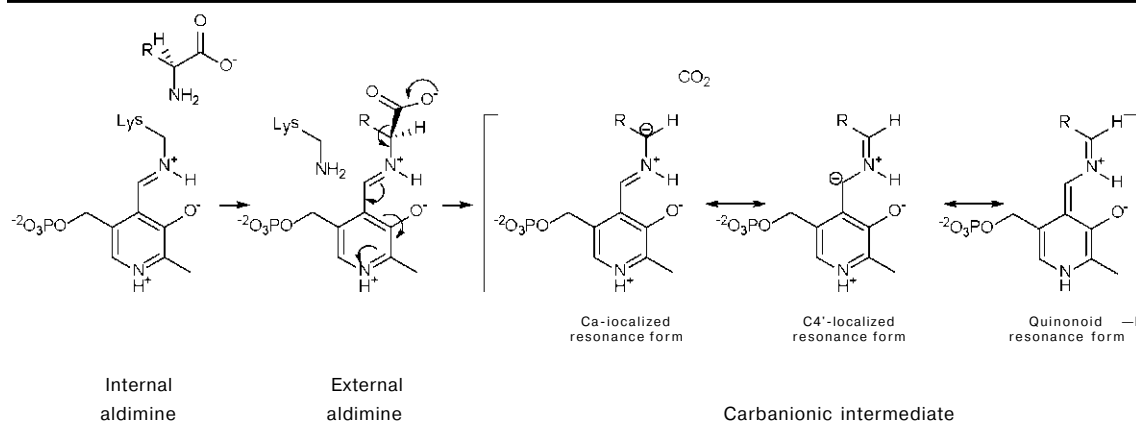
- 1- Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938-945.
- 2- Wolfenden, R. *Chem. Rev.* **2006**, *106*, 3379-3396.
- 3- F. J. W. Roughton *J. Am. Chem. Soc.* **1941**, *63*, 2930-2934.
- 4- Schroeder, G. K.; Lad, C.; Wyman, P.; Williams, N. H.; Wolfenden, R. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4052-4055.
- 5- Major, D. T.; Gao, J. *J. Am. Chem. Soc.* **2006**, *128*, 16345-16357.
- 6- Sinder, M. J.; Wolfenden, R. *J. Am. Chem. Soc.* **2000**, *122*, 11507-11508.
- 7- Walsh, C. In *Enzymatic reaction mechanisms*; Barlett, A. C.; McCombs, L. W. Eds.; W. H. Freeman & Co. New York **1979**; pp 800-801
- 8- Toney, M. D. *Biochim. Biophys. Acta* **2011**, *1814*, 1407-1418.
- 9- Dunathan, H. C. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 712-716.
- 10- Griswold, W. R.; Castro, J. N.; Fisher, A. J.; Toney, M. D. *J. Am. Chem. Soc.* **2012**, *134*, 8436-8438.
- 11- Crueiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. *J. Am. Chem. Soc.* **2011**, *133*, 3173-3183.
- 12- Toth, K.; Richard, J. P. *J. Am. Chem. Soc.* **2007**, *129*, 3013-3021.
- 13- Griswold, W. R.; Toney, M. D. *J. Am. Chem. Soc.* **2011**, *133*, 14823-14830.
- 14- Crueiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 15815-15824.
- 15- Casanovas, R.; Adrover, M.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. *J. Phys. Chem. B* **2012**, *116*, 10665-10675.
- 16- Bach, R. D.; Canepa, C.; Glukhovtsev, M. N. *J. Am. Chem. Soc.* **1999**, *121*, 6542-6555.
- 17- Toney, M. D. *Biochemistry* **2001**, *40*, 1378-1384.
- 18- Limbach, H. H.; Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Shederovich, I. G.; Denisov, G. S.; Toney, M. D. *Biochim. Biophys. Acta* **2011**, *1814*, 1426-1437.
- 19- Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H. H. *Biochemistry* **2010**, *49*, 10818-10830.
- 20- Sharif, S.; Denisov, G. S.; Toney, M. D.; Limbach, H. H. *J. Am. Chem. Soc.* **2007**, *129*, 6313-6327.
- 21- Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, *73*, 383-415.
- 22- Sharif, S.; Fogle, E.; Toney, M. D.; Denisov, G. S.; Shederovich, I. G.; Buntkowsky, G.; Tolstoy, P. M.; Chan-Huot, M.; Limbach, H. H. *J. Am. Chem. Soc.* **2007**, *129*, 9558-9559.
- 23- Lin, Y.; Gao, J. *Biochemistry* **2010**, *49*, 84-94.
- 24- Lin, Y.; Gao, J. *J. Am. Chem. Soc.* **2011**, *133*, 4398-4403.
- 25- Jackson, L. K.; Goldsmith, E. J.; Phillips, M. A. *J. Biol. Chem.* **2003**, *278*, 22037-22043.
- 26- Jackson, L. K.; Brooks, H. B.; Myers, D. P.; Phillips, M. A. *Biochemistry* **2003**, *42*, 2933-2940.
- 27- Fogle, E. J.; Toney, M. D. *Biochim. Biophys. Acta* **2011**, *1814*, 1113-1119.
- 28- Tabor, C. W.; Tabor, H. *Microbiol. Rev.* **1985**, *49*, 81-98.

- 
- 29- Wang, C. C. *Annu. Rev. Pharmacol. Toxicol.* **1995**, *35*, 93-127.
- 30- Marton, L. J.; Pegg, A. E. *Annu. Rev. Pharmacol. Toxicol.* **1995**, *35*, 55-91.
- 31- McCann, P. P., and Pegg, A. E. *Pharm. and Therap.* **1992**, *54*, 195-215
- 32- Dufe, V. T.; Ingner, D.; Heby, O.; Khomutov, A. R.; Persson, L.; Al-Karadaghi, S. *Biochem. J.* **2007**, *405*, 261-268
- 33- a) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede T. *Bioinformatics* **2006**, *22*, 195-201. b) Kiefer, F.; Arnold, K.; Künzli, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Research* **2009**, *37*, D387-D392. c) Peitsch, M. C. *Bio Technology* **1995**, *13*, 658-660.
- 34- Jackson, L. K.; Brooks, H. B.; Osterman, A. L.; Goldsmith, E. J.; Phillips, M. A. *Biochemistry* **2000**, *39*, 11247-11257.
- 35- Wu, Y.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124*, 024503.
- 36- Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712-725.
- 37- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- 38- Ceriotti, M.; Bussi, G.; Parrinello, M. *J. Chem. Theor. Comput.* **2010**, *6*, 1170-1180.
- 39- Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comp. Chem.* **2004**, *25*, 1157 - 1173.
- 40- Wang, J.; Wang, W.; Kollman P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247260.
- 41- Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B*, **2009**, *113*, 6378-96.
- 42- Gaussian 09, Revision B.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.
- 43- a) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984**, *5*, 129-45. b) Besler, B. H.; Merz Jr., K. M.; Kollman, P. A. *J. Comp. Chem.* **1990**, *11*, 431-39.
- 44- Case, D. A.; Darden, T. A.; Cheatham, T. E. III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.;

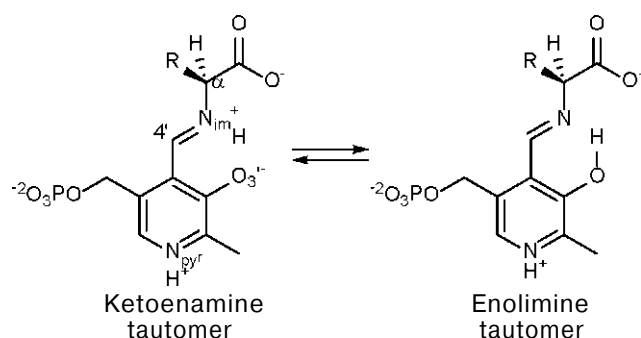


- Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman P. A. **2008**, AMBER 10, University of California, San Francisco.
- 45- Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- 46- Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- 47- CP2K: The CP2K developers group (<http://www.cp2k.org/>) Accessed October **2013**
- 48- Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124–140.
- 49- VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- 50- Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- 51- Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789
- 52- Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703 –1710.
- 53- Blochl, P. E. *J. Chem. Phys.* **1995**, *103*, 7422-7428.
- 54- Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- 55- Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176–1184.
- 56- Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378.
- 57- Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193-4194.
- 58- GLE4MD: <http://gle4md.berlios.de/> Accessed October 2013
- 59- Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- 60- A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562-12566.
- 61- Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533-3539.
- 62- Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- 63- Dapprich, S.; Komáromi, I. n.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.: THEOCHEM* **1999**, *462*, 1-21.
- 64- Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419-1432.
- 65- Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
- 66- Zabinski, R. F.; Toney, M. D. *J. Am. Chem. Soc.* **2001**, *123*, 193-198.
- 67- Brooks, H. B.; Phillips, M. A. *Biochemistry* **1997**, *36*, 15147-15155.
- 68- Jackson, L. K.; Brooks, H. B.; Myers, D. P.; Phillips, M. A. *Biochemistry* **2003**, *42*, 2933-2940.
- 69- Spies, M. A.; Woodward, J. J.; Watnik, M. R.; Toney, M. D. *J. Am. Chem. Soc.* **2004**, *126*, 7464-7475.
- 70- Casasnovas, R.; Salva, A.; Frau, J.; Donoso, J.; Muñoz, F. *Chem. Phys.* **2009**, *355*, 149-156.

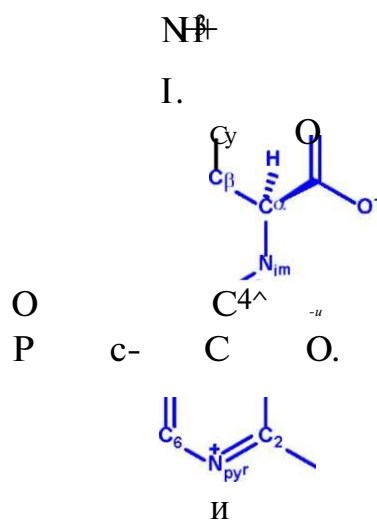
- 71- Lin, Y.; Gao, J.; Rubinstein, A.; Major, D. T. *Biochim. Biophys. Acta* **2011**, *1814*, 1438-1446.
- 72- Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. *J. Chem. Theory Comput.* **2011**, *7*, 1356-1368.
- 73- Oliveira, E. F., Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos *J. Am. Chem. Soc.* **2011**, *133*, 15496-15505.
- 74- Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. *J. Phys. Chem. A* **2004**, *108*, 11709-11714.
- 75- Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. *J. Phys. Chem. A* **2003**, *107*, 9409-9414.



Scheme 1. Transimination reaction between the internal and external aldimines in a PLP-dependent enzyme and decarboxylation of Ca to produce a delocalized carbanionic intermediate.



Scheme 2. Tautomeric equilibrium between the ketoenamine and enolimine forms of PLP Schiffbases.



Scheme 3. QM (blue) and MM (red) regions considered for the simulations in solution and in ODC. All the atoms were treated quantum mechanically in the gas-phase simulations.

Table 1. Calculated and experimental activation free energies and reaction free energies for the decarboxylation of amino acids in different conditions.

		AG*	AGr
		(kcal/mol)	(kcal/mol)
Gly (aq.)	Exptl. <sup>a</sup>	40.0	
PLP-Orn (aq.)	Metadynamics <sup>b</sup>	17.5	10.6
	BLYP <sup>c</sup>	16.6	15.3
	M06-2X <sup>c</sup>	22.6	15.2
PLP-AIB (aq.)	Exptl. (pH 5) <sup>d</sup>	24.8	
PLP-AIB (aq.)	Exptl. (pH 8) <sup>e</sup>	26.6	
PLP-Orn (ODC)	Metadynamics	13.8	11.9
	BLYP	14.2	7.9
	M06-2X	19.6	3.8
	Exptl. <sup>f</sup>	14.5	
PLP-Orn (gas)	Metadynamics	2.4	-8.9
	BLYP	1.4	-14.9
	M062X	1.4	-23.0

<sup>a</sup>Free energy of activation of the uncatalyzed decarboxylation of glycine in solution (Wolfenden2000). Since the metadynamics simulations are performed in the NVT ensemble, the free energy differences correspond to *Helmholtz* free energies instead of Gibbs free energies. <sup>c</sup>The BLYP and M06-2X labels stand for the static calculations performed by using such exchange-correlation functionals as explained in the Methods section. <sup>d</sup>Free energy of activation of the PLP-catalyzed decarboxylation of a-aminoisobutyrate in solution at pH 5 (Toney2000). <sup>e</sup>Free energy of activation of the PLP-catalyzed decarboxylation of a-aminoisobutyrate in solution at pH 8. This value has been calculated by considering the same variation observed for the decarboxylation rate of PLP-2-methyl-2-aminomalonate upon pH increase from 5 to 8 (Toney2000). <sup>f</sup>

Free energy of activation of the PLP-catalyzed decarboxylation of ornithine in the ODC active site (1999Phillips).

Table 2. Relevant average geometrical parameters for the decarboxylation of the PLP-Orn aldimine in solution, ODC active site and gas-phase.<sup>a</sup>

	Aqueous Solution			ODC active site			Gas-phase		
	Equil. <sup>b</sup>	Reac. <sup>c</sup>	Prod. <sup>d</sup>	Equil.	Reac.	Prod.	Equil.	Reac.	Prod.
<b>Carboxylate interactions</b>									
d Ca-Coo <sup>e</sup>	1.58	1.75	3.41	1.65	1.90	3.53	1.69	1.79	3.41
d cOO-N <sub>Asn398</sub> (Ow) <sup>f</sup>	2.77	2.77	3.26	3.33	3.53	4.20			
d cOO-H2NAsn398 (Hw)	1.86	1.79	3.89	2.47	2.66	3.66			
<b>Imine-phenol interactions</b>									
d Nim-H	1.05	1.07	1.21	1.05	1.08	1.39	1.05	1.06	1.59
d O3'-H	1.95	1.78	1.51	1.82	1.71	1.27	1.93	1.76	1.10
d O3'-Ow	2.77	2.77	2.77	2.85	2.92	2.92			
d O3'-Hw	1.86	1.86	1.79	1.91	1.97	2.11			
<b>Pyridine nitrogen interactions</b>									
d N <sub>pyr</sub> -H	1.04	1.04	1.03	1.05	1.05	1.04	1.03	1.03	1.02
d N <sub>pyr</sub> -O <sub>C1274</sub> (Ow)	2.77	2.84	2.84	2.69	2.69	2.72			
d H-OG1u274 (Ow)	1.79	1.79	1.79	1.69	1.69	1.71			
<b>Ca and C4' interactions</b>									
d C4'-N <sub>Ly69</sub>				5.21	5.41	5.51			
d Ca-N <sub>Ly69</sub>				6.34	6.54	6.72			
<sup>d</sup> <sup>cys360</sup> SCys360				6.66	6.07	6.28			
d Ca-SCys360				6.61	6.00	6.17			
<b>Protein-Protein interactions</b>									
<sup>d</sup> <sup>cys360</sup> <sup>asn398</sup> Asn398				2.32	2.52	4.53			
<sup>d</sup> <sup>lys69</sup> <sup>asp88</sup> Asp88				2.71	2.70	2.68			
<b>Hydrogen bonds</b>									
cOO-Ow <sup>g</sup>	2.2	1.7	0.2						
O3'-Ow	1.1	1.1	1.8	0.8	0.8	0.8			
<b>PLP-Orn dihedrals</b>									
6 Coo-Ca-Ni <sub>in</sub> -C4' <sup>h</sup>	-138	-109	-76	98	101	116	-123	-86	-63
6 H-N <sub>in</sub> -C4'-C4	-1	1	0	1	2	3	-1	3	1
6 Nim-C4'-C4-C3	-3	0	-2	0	2	0	1	0	0

<sup>a</sup>The corresponding standard deviation values are given as supporting information material. <sup>b</sup>Averages obtained from the QM or QM/MM equilibrium simulations of the PLP-Orn aldimines. <sup>c</sup>Averages obtained from the QM or QM/MM metadynamics simulations of the configurations corresponding to the reactants state region. <sup>d</sup>Averages obtained from the QM or QM/MM metadynamics simulations of the configurations corresponding to the products state region. <sup>e</sup>All bond lengths and interatomic distances are given in Angstrom. Interatomic distances between the carboxylate oxygens of the PLP-Orn aldimine with the NH<sub>2</sub>- group of Asn398 in the ODC active site or with water molecules in solution. <sup>f</sup>Average number of hydrogen bonds per oxygen atom of the PLP-Orn carboxylate group. The torsion values are given in degrees.

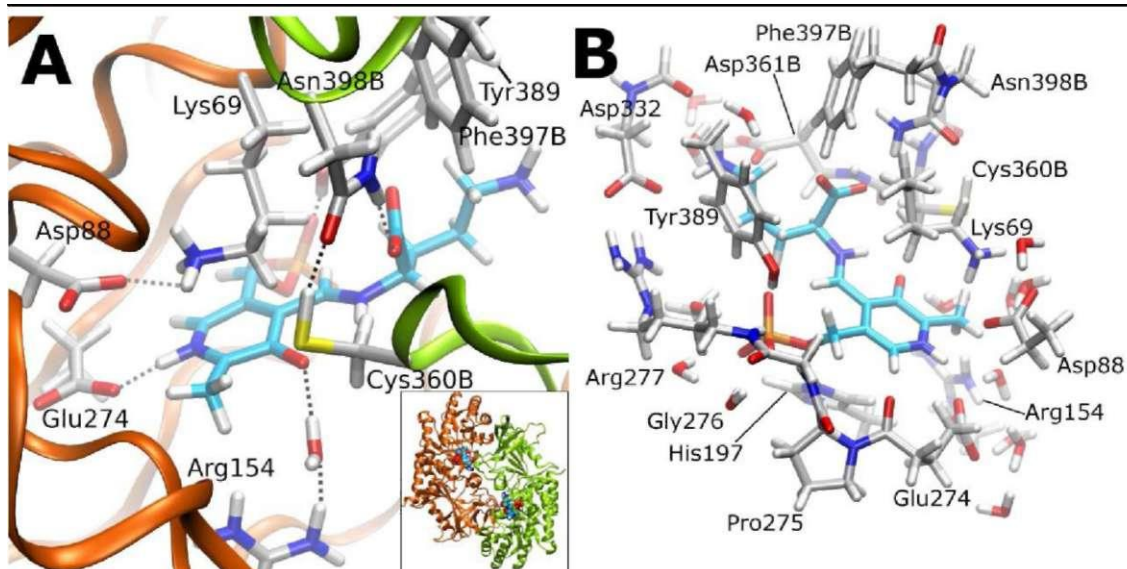


Figure 1. Relevant contacts for the decarboxylation reaction of the PLP-Ornithine aldimine in the ODC active site (A). Ornithine decarboxylase enzyme showing the two active sites formed at the dimer interface (Inset). Cluster model of the ODC active site used in the ONIOM static calculations (B).

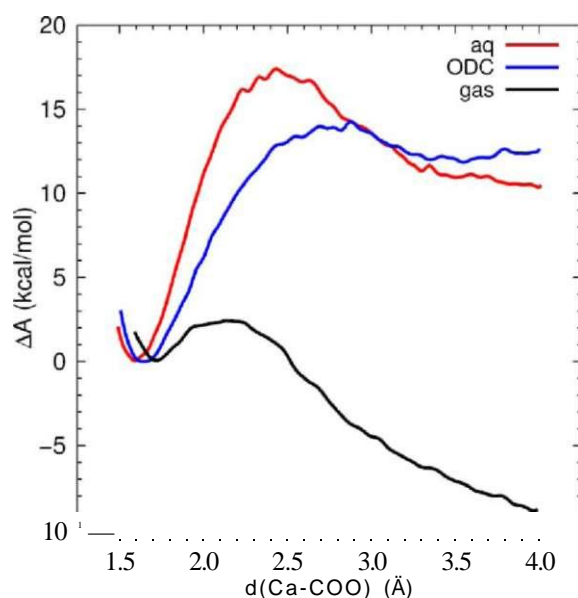


Figure 2. Free energy profiles obtained from the metadynamics simulations for the PLP-catalyzed decarboxylation of ornithine in aqueous solution (red), ODC active site (blue) and gas-phase (black).

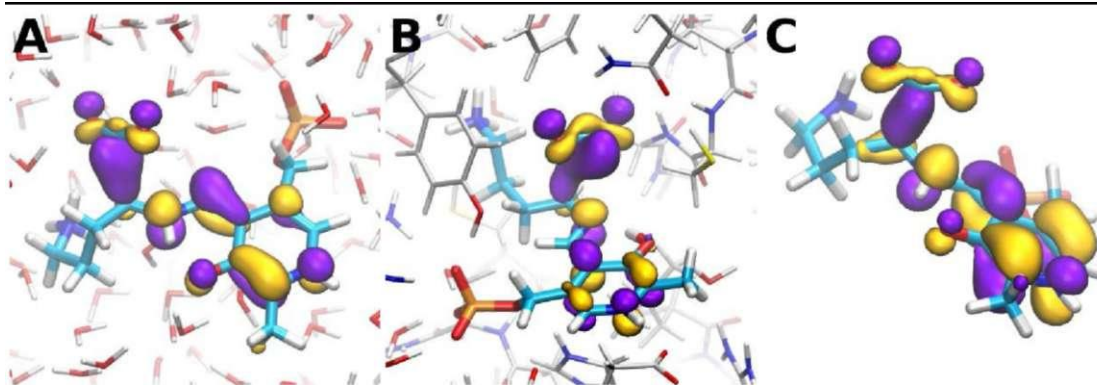


Figure 3. Wannier maximally localized functions corresponding to the HOMO-1 of the decarboxylation transition states obtained from the committor analysis in aqueous solution (A), ODC active site (B) and gas-phase (C). In all cases, the electron pair of the Ca-COO a bond is also delocalized across the imine-pyridine n system.

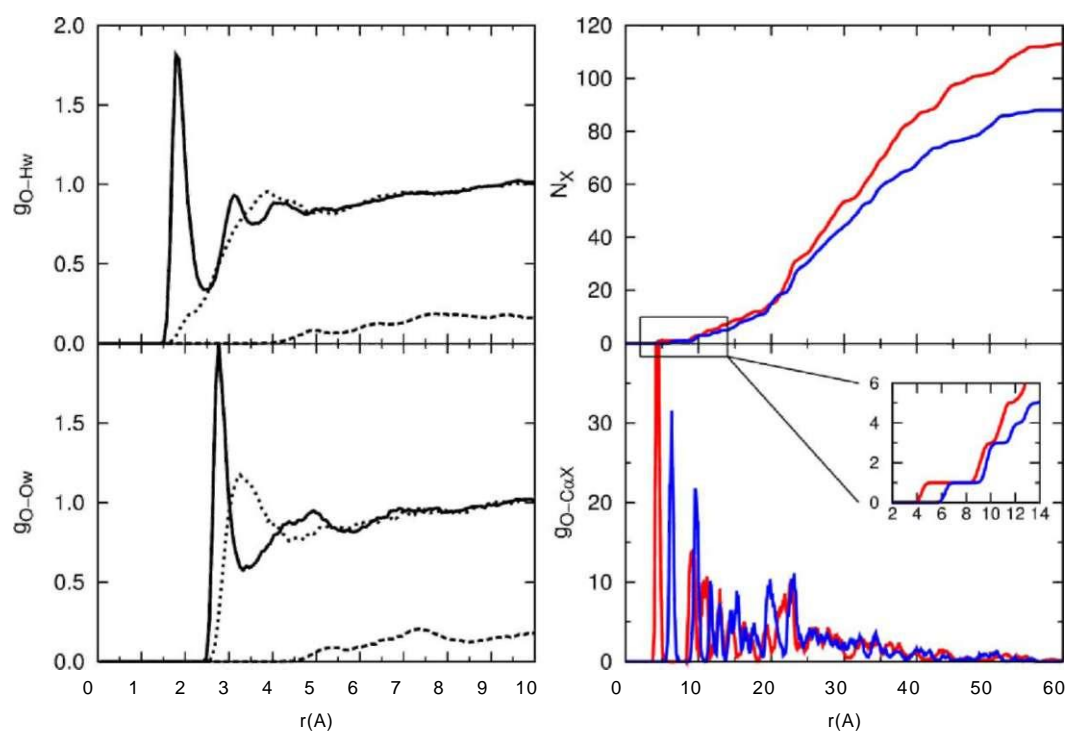


Figure 4. Radial distribution functions of the carboxylate oxygens and water hydrogens (top left) or water oxygens (bottom left) for the reactants (solid lines) and products (dotted lines) in aqueous solution. The respective radial distribution functions for the reactants state in the ODC active site (dashed lines) show that the enzyme environment disfavors the reactants state. The bottom right plots correspond to the radial distribution functions of the carboxylate oxygens and the alpha carbons of the anionic residues (red line) or cationic residues (blue line) in the ODC enzyme. The top right plots depict the integrated number of anionic (red line) and cationic (blue line) residues as a function of the distance from the carboxylate group in the ODC enzyme. The inset depicts such integrated number of anionic and cationic residues in the ODC active site (i.e. the first 14Å away from the carboxylate group).



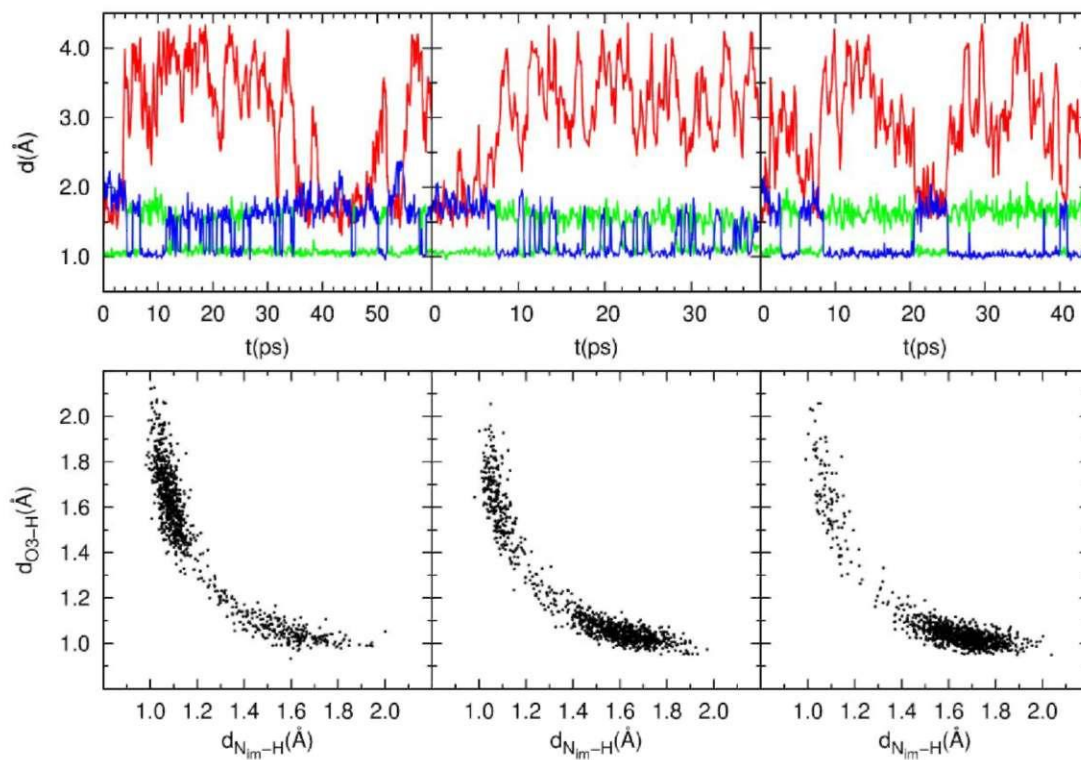


Figure 5. (Top) Evolution of the Ca-COO (red lines), O3'-H (blue) and  $N_{im}$ -H (green) interatomic distances during the metadynamics simulations in solution (left), ODC active site (center) and gas-phase (right). (Bottom) Distribution of the proton location in the  $N_{im}$ -O3' intramolecular hydrogen bond for the products state (i.e. carbanionic intermediates) in solution (left), ODC active site (center) and gas-phase (right).



**Extraordinaire decarboxylation rates catalyzed by modestly efficient enzymes. A QM/MM metadynamics study on the enzymatic and nonenzymatic pyridoxal 5'-phosphate-catalyzed decarboxylation of amino acids**

Rodrigo Casasnovas, Juan Frau, Sebastiano Caravati, Josefa Donoso, Francisco Muñoz, Michele Parrinello

Supporting Information

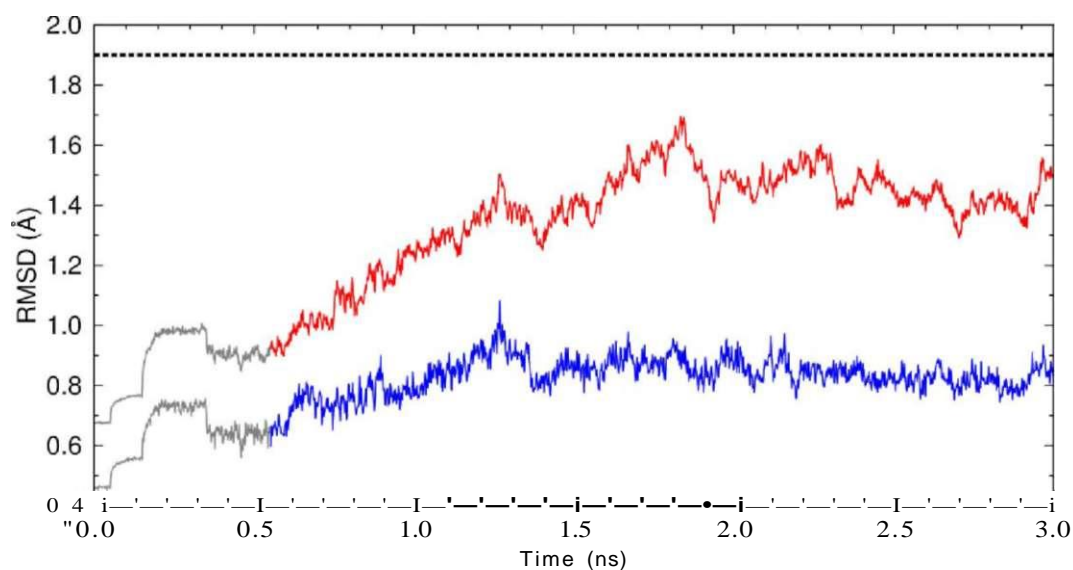


Figure S1. Root Mean Square Deviation (RMSD) values of the protein backbone (red) and of the active site residues (blue) during the classical molecular dynamics equilibration simulation. The grey fractions of the plots correspond to the initial stage of the protein equilibration in which harmonic constraints were applied. The RMSD values were calculated with respect to the crystallographic structure PDB 2000. The resolution of the X-ray experiment (i.e. 1.9 Å) is indicated as a dotted horizontal line.

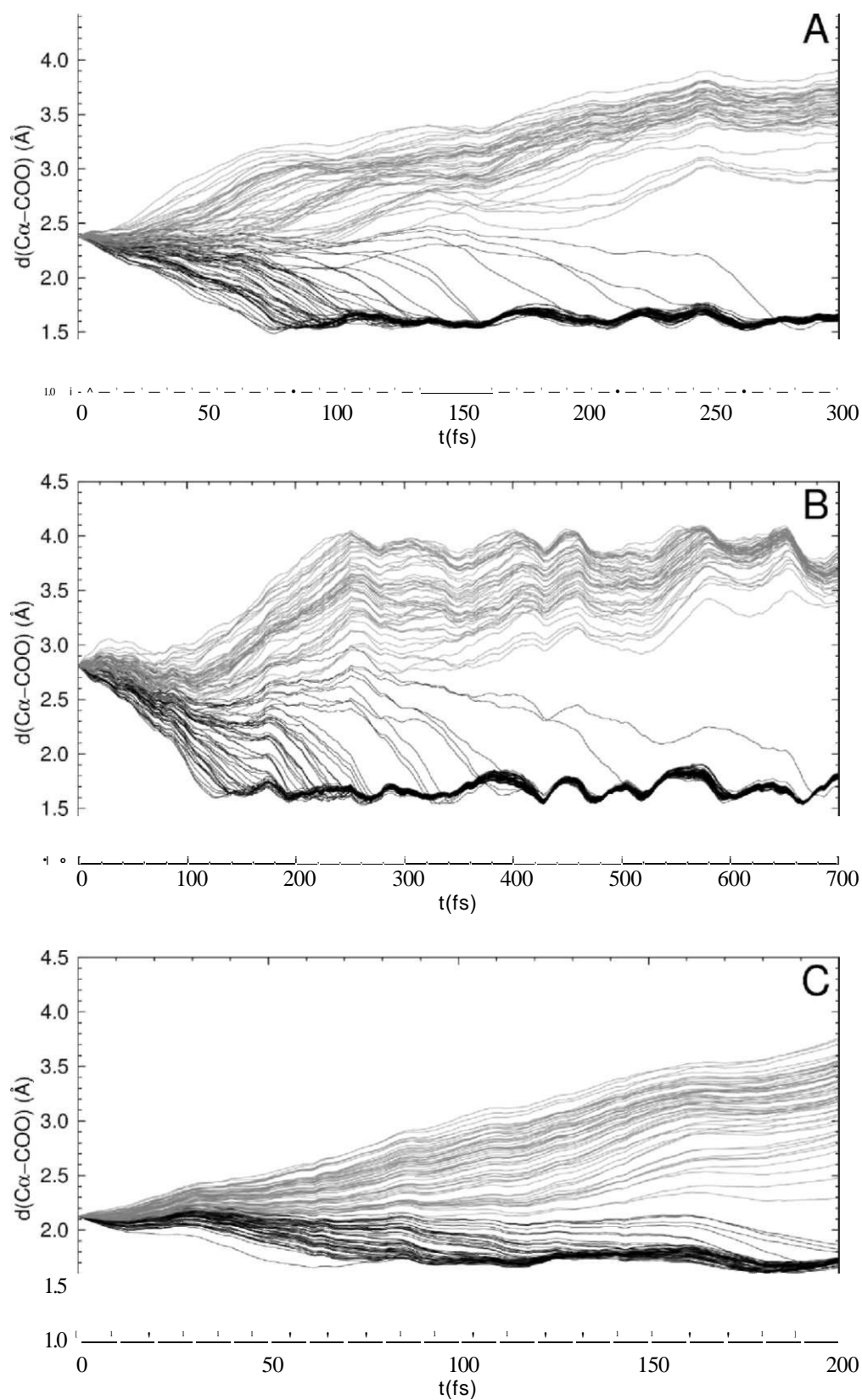


Figure S2. Evolution of the Ca-COO bond length in the committor analysis of the decarboxylation in aqueous solution (A), ODC active site (B) and gas-phase (C). The trajectories that lead to reactants and products are respectively represented by black and grey lines.

Figure S3-1

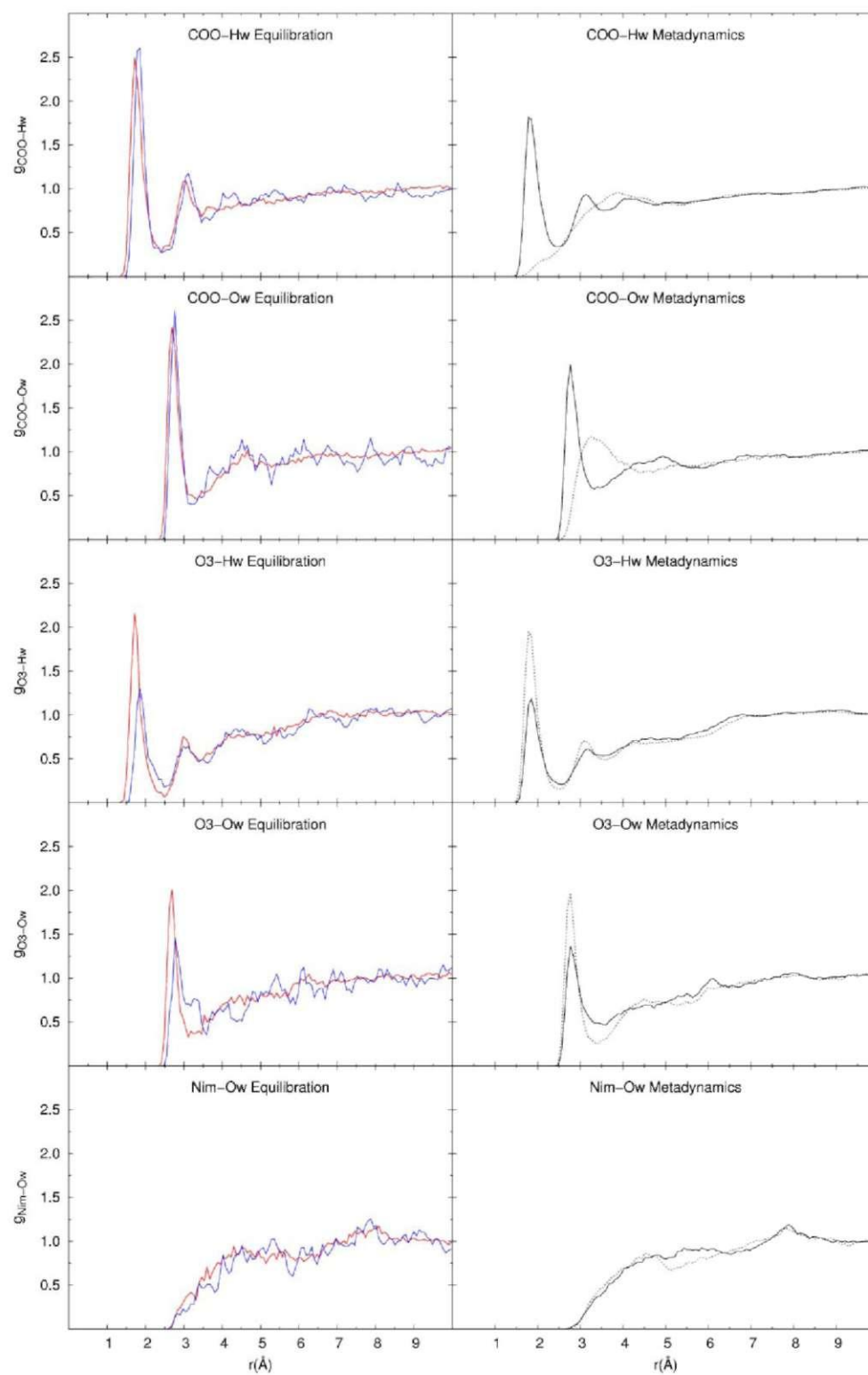


Figure S3-2

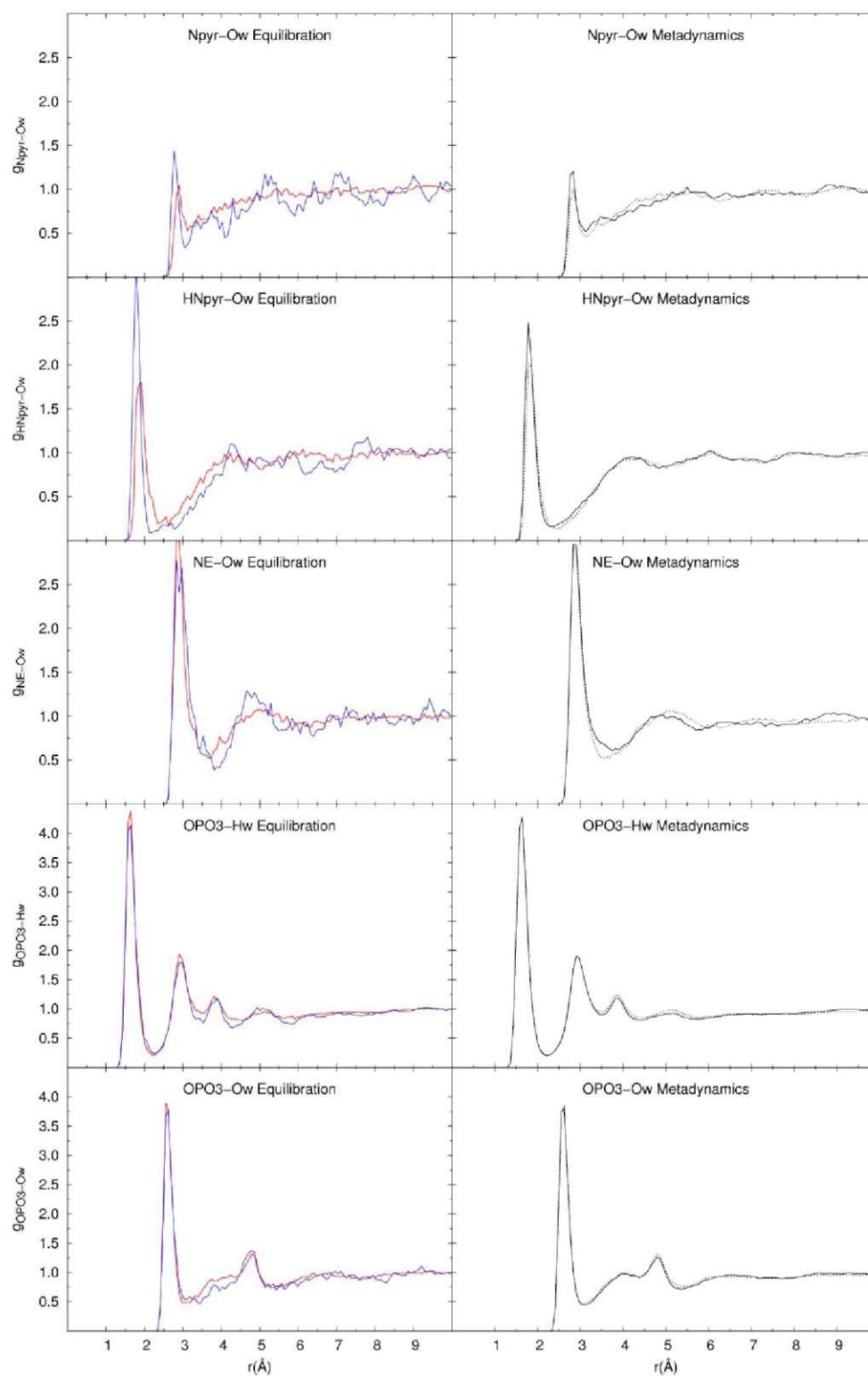


Figure S3. Radial distribution functions of selected atoms in the PLP-Orn aldimine and water hydrogens (Hw) or water oxygens (Ow). The plots in the left-hand side compare the radial

distribution functions of the MM (red solid lines) and the QM/MM (blue solid lines) equilibration simulations. The plots in the right-hand side correspond to the radial distribution functions calculated from the metadynamics simulations independently for the reactants state (black solid lines) and the products state (black dotted lines). The label COO stands for the carboxylate oxygens, O3 for the phenol oxygen, Nim for the imine nitrogen, Npyr for the pyridine nitrogen, HNpyr for the proton bound to the pyridine nitrogen, NE for the nitrogen atom of the  $\epsilon$ -amino group, and OPO3 for the phosphate oxygens.

Figure S4. Evolution of the Ca-COO (red lines), O3-H (green lines) and Nim-H (blue lines) bond lengths for each "walker" during the metadynamics simulations in solution (figure S4-1), in gas-phase (figure S4-2) and in the ODC active site (figure S4-3).

Figure S4-1

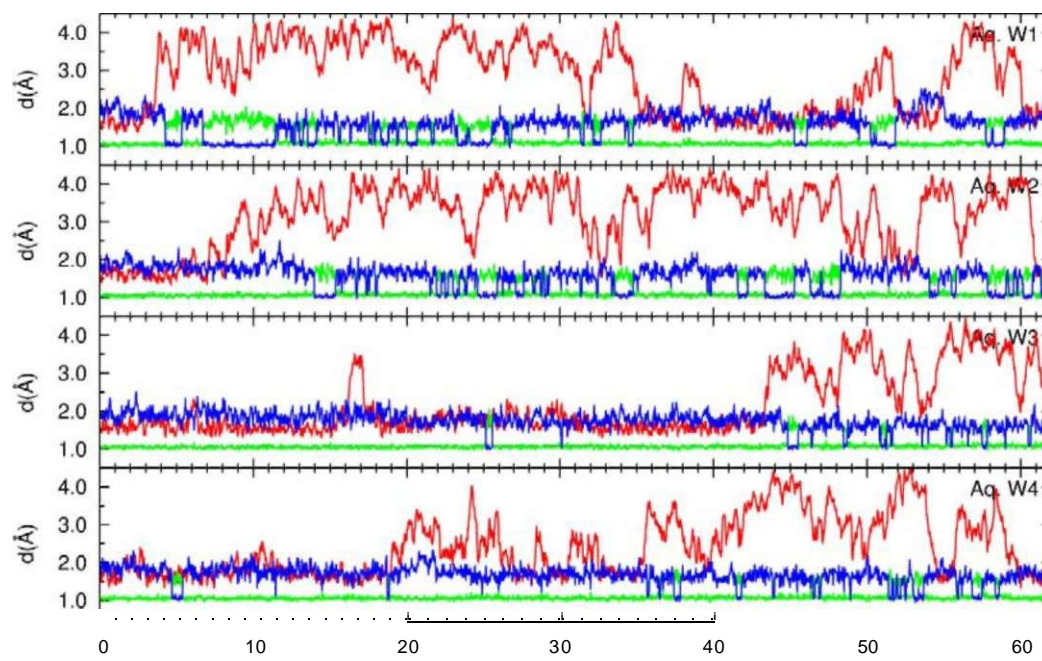


Figure S4-2

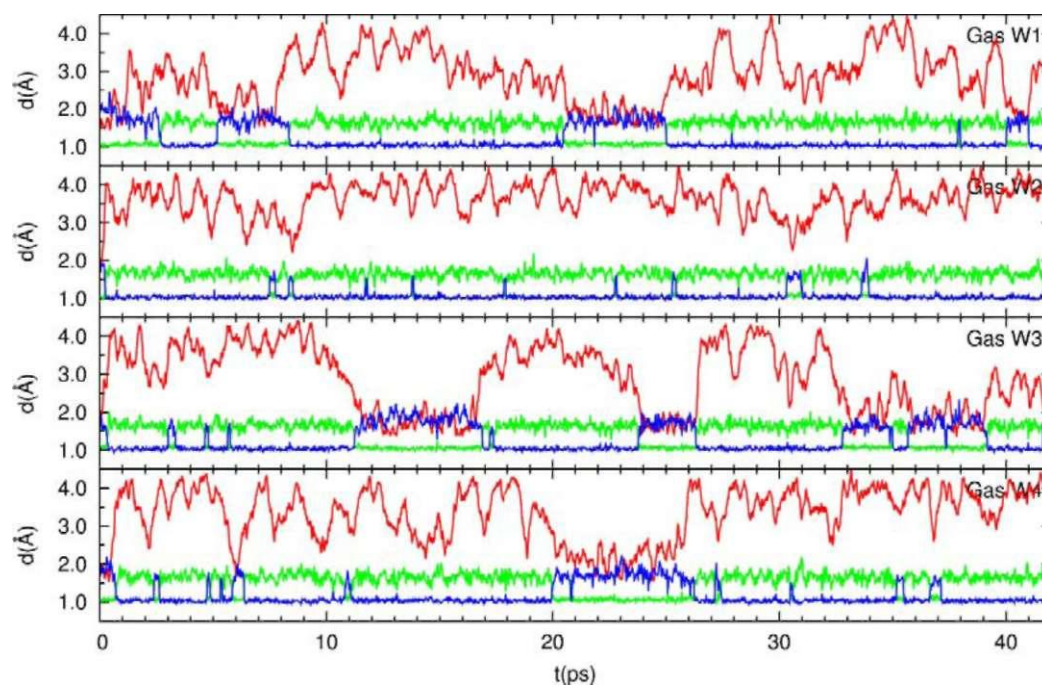




Figure S4-3

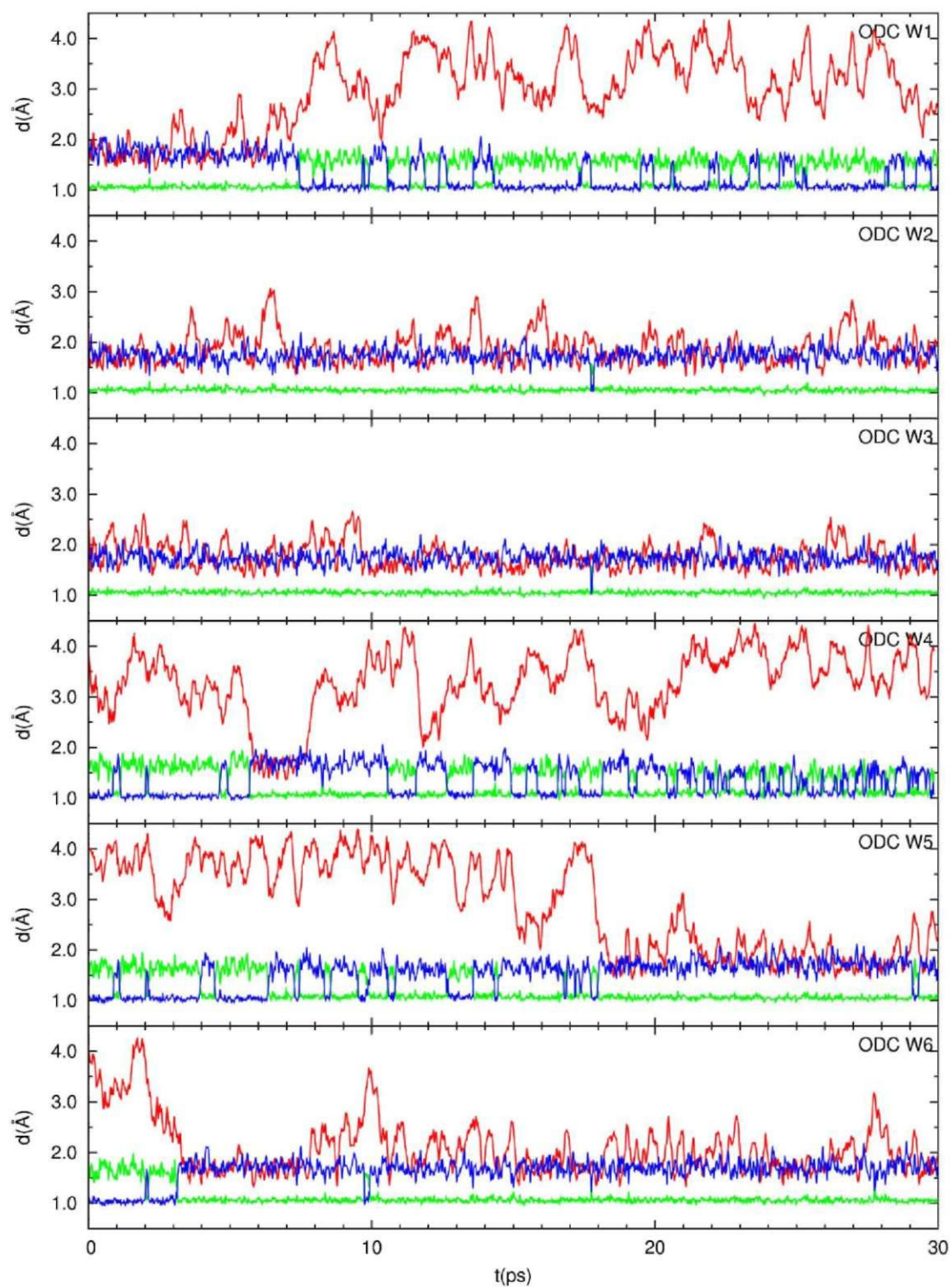


Table S1. Standard deviation values of the averages reported in the article.

	Aqueous Solution			ODC active site			Gas-phase		
	Equil.	Reac.	Prod.	Equil.	Reac.	Prod.	Equil.	Reac.	Prod.
<b>Carboxylate interactions</b>									
d Ca-Coo	0.04	0.2	0.6	0.06	0.4	0.4	0.08	0.19	0.6
d cOO-NAsn398(Ow) <sup>a</sup>	0.14	0.14	0.14	0.4	0.4	0.4			
d cOO-H2NAsn398(Hw)	0.14	0.14	0.14	0.5	0.5	0.4			
<b>Imine-phenol interactions</b>									
d Nim-H	0.03	0.10	0.2	0.03	0.11	0.3	0.03	0.08	0.2
d O3'-H	0.2	0.2	0.3	0.14	0.19	0.3	0.18	0.18	0.19
d O3'-Ow <sup>a</sup>	0.14	0.14	0.14	0.13	0.18	0.2			
d O3'-Hw <sup>a</sup>	0.14	0.14	0.14	0.2	0.3	0.5			
<b>Pyridine nitrogen</b>									
d Npyr-H	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
d Npyr-OGlu274 (Ow) <sup>a</sup>	0.14	0.14	0.14	0.07	0.07	0.08			
d H-OGlu274 (Ow) <sup>a</sup>	0.14	0.14	0.14	0.08	0.08	0.09			
<b>Ca and C4' interactions</b>									
d C4'-NLys69				0.2	0.3	0.3			
d Ca-NLys69				0.2	0.3	0.3			
<sup>a</sup> Cys360				0.3	0.5	0.8			
<sup>a</sup> Cys360				0.3	0.5	0.9			
<b>Protein-Protein interactions</b>									
<sup>a</sup> Cys360 <sup>a</sup> Asn398				0.6	1.0	0.9			
<sup>a</sup> Lys69 <sup>a</sup> Asp88				0.08	0.09	0.08			
<b>Hydrogen bonds</b>									
cOO-Ow	0.8	0.9	0.5						
O3'-Ow	0.7	0.8	0.8	0.4	0.4	0.4			
<b>PLP-Orn dihedrals</b>									
9 Coo-Ca-Nim-C4'	15	25	29	98	10	12	16	25	35
9 H-Nim-C4'-C4	9	9	10	9	8	9	10	8	9
9 Nim-C4'-C4-C3	10	9	9	8	8	8	8	8	9

<sup>a</sup>The bin used in the construction of the histograms of the radial distribution functions was 0.07Å. A standard deviation double of the bin size is reported for the values obtained from the radial distribution functions.







## **5. Discussion**



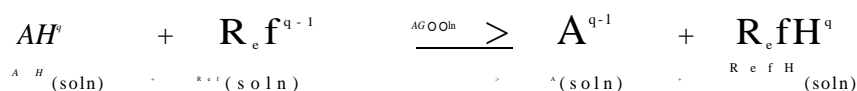
### 5.1. Theoretical $pK_a$ calculations

The precision of the theoretical  $pK_a$  determination via thermodynamic cycles depends on the errors in the gas-phase dissociation free energy of the studied acid and in the solvation free energies of the involved species. Since the best *ab initio* and DFT methods provide reaction energies with a precision of  $\sim 1$  kcal/mol, if all errors were additive, the cumulative error throughout the cycle would be at least within 3-5 kcal/mol (i.e. 2.2-3.7  $pK_a$  units). In fact, as shown in the results section 4.1.1., it is possible to obtain gas-phase deprotonation free energies with average errors of  $\sim 1$  kcal/mol at affordable computational costs for medium-sized molecules. However, the  $pK_a$  calculations of pyridines by using thermodynamic cycle 1 (Scheme 19) show average absolute errors of 2.3  $pK_a$  units (Results 4.1.2).

Addition of a single explicit solvent water molecule interacting with the pyridine nitrogen in the calculation of the solvation free energies significantly improves the  $pK_a$  predictions (Results 4.1.2). Since the errors obtained when considering explicit water molecules are lower than the expected 2.2-3.7  $pK_a$  units, it should be concluded that the improvement of the predictions is due to better cancelation of errors. The combination of continuum models and methods/basis sets that provide the lowest errors coincide in all the three considered thermodynamic cycles (Results 4.1.2). It should be noted that the errors of the calculated free energies cancel out in different extent for each thermodynamic cycle because their associated equations for the  $pK_a$  calculation differ. Therefore, the improvement in the  $pK_a$  predictions is also due to the reduction of errors in the individual solvation energies by the inclusion of explicit solvent molecules.

Ionic and zwitterionic species exhibit, in absolute values, larger solvation energies and larger associated errors than neutral species. An analysis of the solvation cavities reveals that a significant fraction of the total electrostatic solute-solvent interaction is calculated over the spheres of the more charged atoms (Results 4.1.2). Therefore, when the more charged atoms are less exposed to the solvent, slight variations in the area of their spheres entail significant variations in the total solvation energy. In these cases the introduction of explicit solvent molecules allows a more realistic description of the short-range solute solvent interactions. In addition, the explicit solvent molecules that interact with the more ionic/polar groups are polarized and since they are much more exposed to the continuum solvent, the area over which the main electrostatic terms are calculated is much larger, which reduces the errors in the solvation energies (Results 4.1.2).

Given that the evaluation of thermodynamic cycles indicates that accurate  $pK_a$  predictions require maximizing the cancelation of errors, which are mostly associated to the solvation energies, an isodesmic reaction in solution is proposed as an alternative strategy (Scheme 22). An isodesmic reaction is a chemical reaction in which the number and type of chemical bonds in the products and reactants sides are equal.



Scheme 22. Isodesmic reaction consisting on a proton exchange reaction between an acid species AH of global charge  $q$  and a reference species RefH of global charge  $q'$ .

Opposite to the thermodynamic cycles, the isodesmic reaction is constituted by a single chemical reaction, which entails that the expected error in the  $pK_a$  prediction is reduced to that of a unique reaction energy value. Considering again that the best quantum chemistry methods provide errors of  $\sim 1$  kcal/mol for reaction energies, the expected precision in the  $pK_a$  predictions is  $\sim 0.7$  units for the isodesmic reaction. However, it is not possible to neglect that the calculated free energies in solution may include part of the errors present in the solvation energies, which have their origin in the approximations of the continuum solvent model. This is the reason why the proposed reaction in solution is isodesmic, to maximize the cancelation of errors in the calculated reaction energy by choosing a reference species that is "similar" to the studied acid.

The free energy of proton exchange is calculated according to equation [90] by considering the absolute free energies in solution of each species involved in the reaction. The calculated free energy of the reaction is nothing but the relative acidity of AH with respect to RefH (Scheme 22), which is directly related to their  $pK_a$  difference. Accordingly, the  $pK_a$  of the species AH is obtained from equation [91], in which the experimental  $pK_a$  value of the reference species is considered.

$$\text{AG}_{\text{ra}}^{\text{m}} = \text{G}^{\text{som}}(\text{AT}^1) + \text{G}^{\text{som}}(\text{RefH}^{q'}) - \text{G}^{\text{A}}(\text{AH}^q) - \text{G}^{\text{R e f H}^{q-1}} \quad [90]$$

$$\text{AG}_{\text{ra}}^{\text{m}} = \text{G}^{\text{som}}(\text{AT}^1) + \text{G}^{\text{som}}(\text{RefH}^{q'}) - \text{G}^{\text{A}}(\text{AH}^q) - \text{G}^{\text{R e f H}^{q-1}} \quad [91]$$

Unlike gas-phase, the calculation of absolute free energies in solution is more problematic from a formal standpoint. The harmonic and rigid rotor approximations constitute an acceptable description of the physics of nuclear thermal motion in gas-phase but not in solution, especially when considering translation and rotation motions. Actually, a much more correct description of the nuclear motions would require simulation techniques like Monte-Carlo or Molecular Dynamics apart from the inclusion of nuclear quantum effects. Furthermore, in order to describe short and long range solute solvent interactions, the solvent molecules should be explicitly treated along with periodic boundary conditions. However, the computational cost of the simulation techniques is contrary to the philosophy of quick and simple  $pK_a$  predictions subjacent to the isodesmic reaction approach.

Several assumptions are made when calculating reaction free energies in aqueous solution according to the isodesmic reaction schema as a compromise between the methodological rigor and the computational cost and simplicity. Firstly, it is assumed that the contributions to the free energy resulting from translational and rotational motions of an acid AH will mostly cancel out with those of its conjugate base

A, since their structures only differ in a proton. Secondly, the harmonic approximation is considered valid to describe the vibrational motions of the solute nuclei. Finally, the possible errors due to the absence of explicit solvent interacting with the solute will cancel out between the species present in both sides of the chemical equation.

The obtained results (4.1.3) point out that the  $pK_a$  values predicted with the isodesmic reaction show an equivalent precision to those predicted with thermodynamic cycles in the most accurate cases. However, the presence of explicit water molecules is not necessary to obtain mean absolute deviation errors (MAD) of  $\sim 1$   $pK_a$  unit, even though their addition reduces the MAD values by  $\sim 0.4$   $pK_a$  units. More importantly, low MAD errors are obtained when the global charges of the studied acid AH and the reference acid species BH are different, while in the case of thermodynamic cycles the global charges strongly determine the magnitude of the errors of the calculated solvation energies and  $pK_a$  values. In addition, the  $pK_a$  predictions with Density Functional Theory methods yield results with equivalent precision than the much more computationally demanding post Hartree-Fock methods. Finally, the absence of gas-phase calculations avoids difficulties related with conformational changes associated to the solvation process and protonation states which are unstable in absence of solvent.

The performance of the isodesmic reaction in the  $pK_a$  calculations lies in a good cancellation of the errors originated mostly in the interactions between the solutes and the continuum solvent model. Taking into account that the total interaction solute-continuum is partitioned over the solvation spheres of the heavy atoms, two forms of error cancellation are distinguished. Firstly, the errors in the interaction of the group that donates the proton of AH cancel out with those of the group of the reference species RefH that accepts it, and likewise for the deprotonated groups of the conjugate bases A and B. Secondly, the errors in the solute-continuum interactions of those heavy atoms that do not modify their protonation state in the reaction cancel out between the conjugate acid base pairs AH/A and RefH/Ref. Therefore, it follows that the chemical structures of the studied acid A and the reference species RefH should be similar to maximize the cancellation of errors. In the ideal case the reference species is identical to the studied acid species, which ensures a perfect cancellation of errors in the solute-continuum interaction. Considering the equations of the isodesmic reaction, the errors in the  $pK_a$  predictions are expected to be lower when the free energy of proton exchange between the acid AH and the reference species RefH approaches to zero.

The plot of the  $pK_a$  values predicted for the carbon acids against the experimental  $pK_a$  values shows that a correct linear trend is predicted in all the  $pK_a$  range (Figure 1). However the errors of the predicted  $pK_a$  values increase with the  $pK_a$  difference between the acid AH and the reference species (Figure 1). It is interesting to note that although two species with similar  $pK_a$  values usually exhibit similar chemical structure and global charge, the case of carbon acids exemplifies that satisfactory  $pK_a$  predictions are made for structures of diverse chemical structures and global charges (Figure 1, Results 4.1.3, 4.2.2, 4.2.3). What is more, for some acids that are structurally similar or show identical global charge than the reference species, higher errors are obtained than for other acids that show a  $pK_a$  value more similar to the reference species (Results 4.1.3, 4.2.2, 4.2.3).

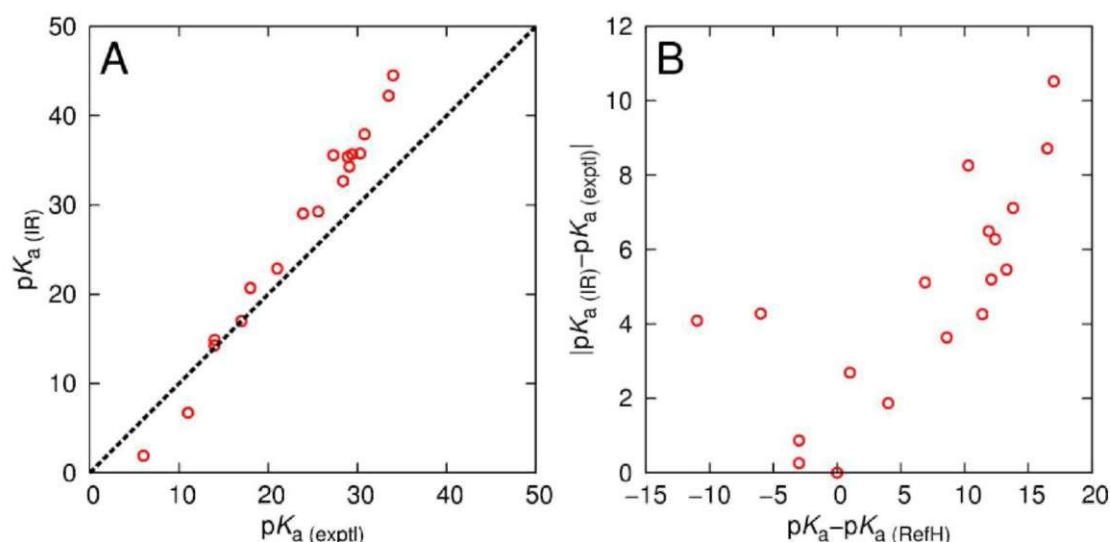


Figure 1. (A) Plot of the predicted  $pK_a$  values for carbon acids with the isodesmic reaction against the corresponding experimental  $pK_a$  values. The dashed line marks the ideal behavior of the predicted  $pK_a$  values. (B) Plot of the absolute errors of the predicted  $pK_a$  values against the  $pK_a$  difference with respect to the reference species.

The alternative case is exemplified by the pyridines (Results 4.1.3), which show identical global charge and very similar chemical structure than the reference species. An analysis of the errors reveals that the largest errors correspond to those species whose  $pK_a$  diverge most from that of the reference species (Results 4.1.3). Taking everything into account, it can be concluded that the isodesmic reaction approach provides the best results when the  $pK_a$  values of the studied acid and the reference species are similar.

At this point, several questions arise. Firstly, how similar should be the  $pK_a$  of the studied acid and the reference species? Which leads to a more interesting point regarding the possibilities of making theoretical predictions: How does the error of the predictions depend on the  $pK_a$  difference between the studied acid and the reference species? However, this is problematic from a practical point of view since the reference species should be chosen on the basis of the similitude between its  $pK_a$  value and that of the studied acid, which is unknown. A possible solution to circumvent this paradox would be to choose several reference species whose  $pK_a$  values span in a certain range, calculate the  $pK_a$  of the studied acid by using all such reference species and decide which one is optimal *a posteriori*. Still, this is not the best solution when multiple predictions have to be made to determine the relative acidities of a set of molecules with unknown  $pK_a$  values.

Instead of employing multiple reference species, an alternative procedure consists on considering their experimental  $pK_a$  values to obtain a unique correction of the free energies calculated with the isodesmic reaction in a large  $pK_a$  range. In this procedure, only one molecule is chosen as the reference species and the free energies of proton exchange of the remaining molecules are calculated following the isodesmic reaction (Scheme 22). Then, considering the relationship between the  $pK_a$  of AH and the calculated free energy given by equation [91], a linear regression is carried out to



obtain the parameters that fit such calculated free energies to the experimental  $pK_a$  values. Finally, such equation is used to predict the acidities of acid species AH of unknown  $pK_a$  by correcting their calculated free energies of proton exchange with respect to the reference species.

The case of the carbon acids is used to exemplify this procedure. As shown in Figure 1, there is a direct proportionality between the predicted and the experimental  $pK_a$  values of carbon acids. However, there is also a deviation from the ideal behavior and the calculated  $pK_a$ 's are overestimated in absolute value. The linear regression of the experimental  $pK_a$  values with the calculated free energies divided by the  $RT\ln 10$  factor results in a line of good coefficient of determination ( $R = 0.988$ , Results 4.2.2, 4.2.3) (Figure 2). The resulting slope and y-interception are 0.67 and 16.56, which according to equation [91] should be respectively 1 and 17 (i.e. the  $pK_a$  of the reference species). These regression parameters confirm that the errors of the predicted  $pK_a$  values with the isodesmic reaction are directly proportional to the magnitude of the calculated free energy of proton exchange, which also involves that the best predictions are obtained when the  $pK_a$  of the acid AH is similar to that of the reference species.

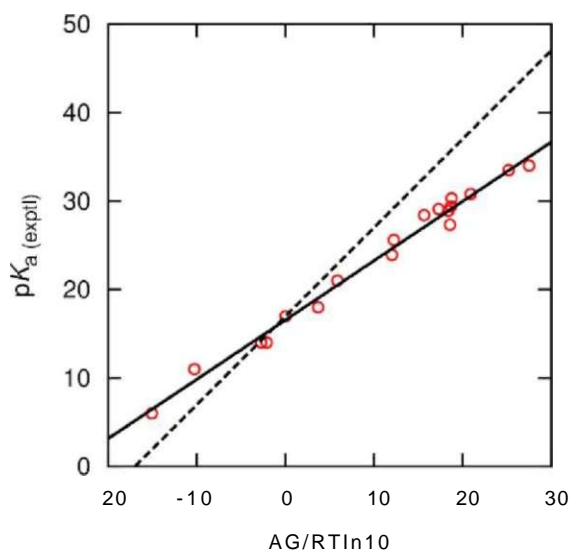


Figure 2. Linear regression of the experimental  $pK_a$  values of the carbon acids and their calculated free energies of proton exchange according to the isodesmic reaction. The result of the regression is shown as a solid line of slope 0.67 and y-intercept 16.56. The ideal behavior is shown as a reference by a dashed line of slope 1 and y-intercept 17.

Slope values lower than 1 involve that the calculated free energies according to the isodesmic reaction scheme are overestimated in absolute value, which results in the overestimation of the  $pK_a$  values shown in Figure 1. Large  $pK_a$  differences of the same type of functional group in two molecules are usually related to differences in their global charge. Accordingly, the overestimation in the energies of proton exchange is a manifestation of the limitations of continuum solvent models when calculating the interaction with ionic solutes, which is also the main source of errors in the solvation

energies. However, our results show that correct predictions are made for acids that show different global charge than the reference but similar  $pK_a$  values (4.1.3, 4.2.2, 4.2.3). Therefore, the errors of the calculated free energies should be attributed to differences in the charge distribution and polarity close to the acidic groups of AH and RefH involved in the protonation and deprotonation reaction.

The magnitude of the absolute errors of the  $pK_a$  values calculated with the isodesmic reaction is directly proportional to the  $pK_a$  difference between the acid AH and the reference species RefH by a factor of  $|(1-a)/a|$  (Figure 3B), being  $a$  the slope obtained in the regression. Accordingly, the closer the slope is to unity, the lower are the errors in the predictions. The inverse (i.e.  $|a/(1-a)|$ ) indicates the maximum  $pK_a$  difference between AH and RefH that provides predictions with errors below a certain limit. For example in the case of carbon acids  $a = 0.67$  so, the errors in the predicted  $pK_a$  values will remain below 1 unit on condition that the  $pK_a$  of AH and RefH differ by a maximum of  $\pm 2$  units. However, if the calculated free energies are corrected according to the linear regression, the standard error of the predicted  $pK_a$  values remains constant at 0.8 units in the entire  $pK_a$  range (Figure 3B). It is worth to note that errors of 0.8  $pK_a$  units are significant for common acid functionalities but are within the typical experimental uncertainties of carbon acids.

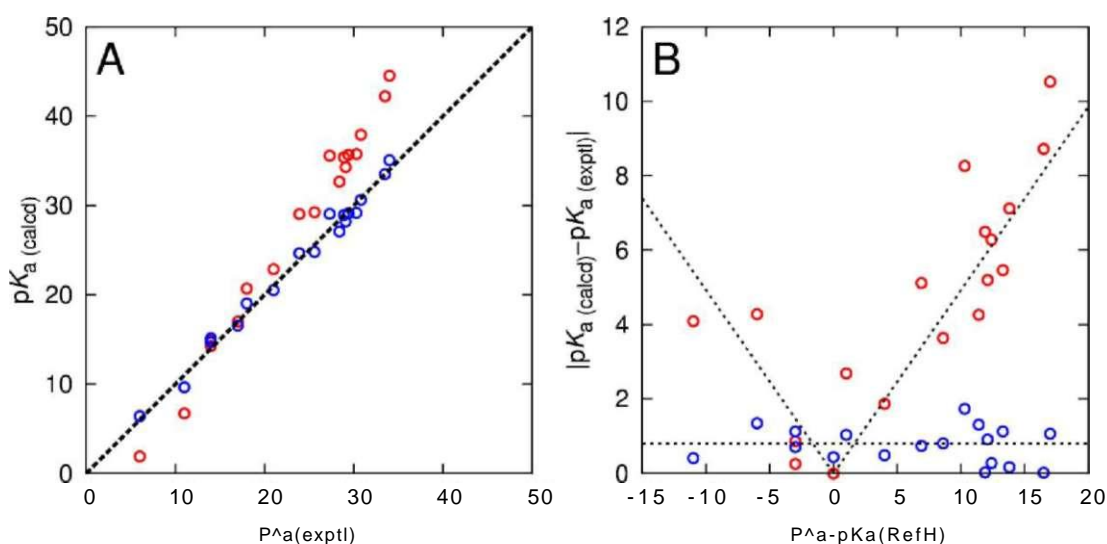


Figure 3. (A) Plot of the predicted  $pK_a$  values with the isodesmic reaction for the carbon acids against their experimental values. The red and blue circles respectively correspond to the predictions made with the isodesmic reaction before and after the corrections to the calculated energies obtained in the linear regression with the experimental values. The ideal behavior is shown as a reference by the dashed line. (B) Plot of the absolute errors of the predicted  $pK_a$  values against the  $pK_a$  difference with respect to the reference species. The dotted lines correspond to the estimated errors of the uncorrected predictions, which depend on the  $pK_a$  difference by a factor of  $|(1-a)/a|$ , and to the corrected predictions, which are constant.

The application of the isodesmic reaction to other acid groups such as pyridines, phenols and benzylamines shows similar behavior than the carbon acids. In all the three

sets of molecules the errors are low if the  $pK_a$  values of the acids AH are close to those of the respective reference species, but otherwise the predicted  $pK_a$ 's are overestimated (Figure 4). The linear regressions of the experimental  $pK_a$  values with the calculated free energies for each set of molecules gives slope values of 0.56, 0.52 and 0.32 respectively for the pyridines, phenols and benzylamines. Therefore, the overestimation of the free energies of proton exchange is larger than in the case of carbon acids, which also suggests that the modeling of the solute solvent interaction by the continuum is worse, especially for the benzylamines. The pyridine, phenol and amine groups are "harder" than carbon acids, which entails that the polar and electrostatic contributions of the solvent interaction are more important and explains the poorer performance of the continuum solvent models. Nevertheless, when the calculated free energies of proton exchange are corrected by considering the regression parameters, in all three cases the standard errors of the predictions are only 0.3 units in the studied  $pK_a$  range (Figure 4).

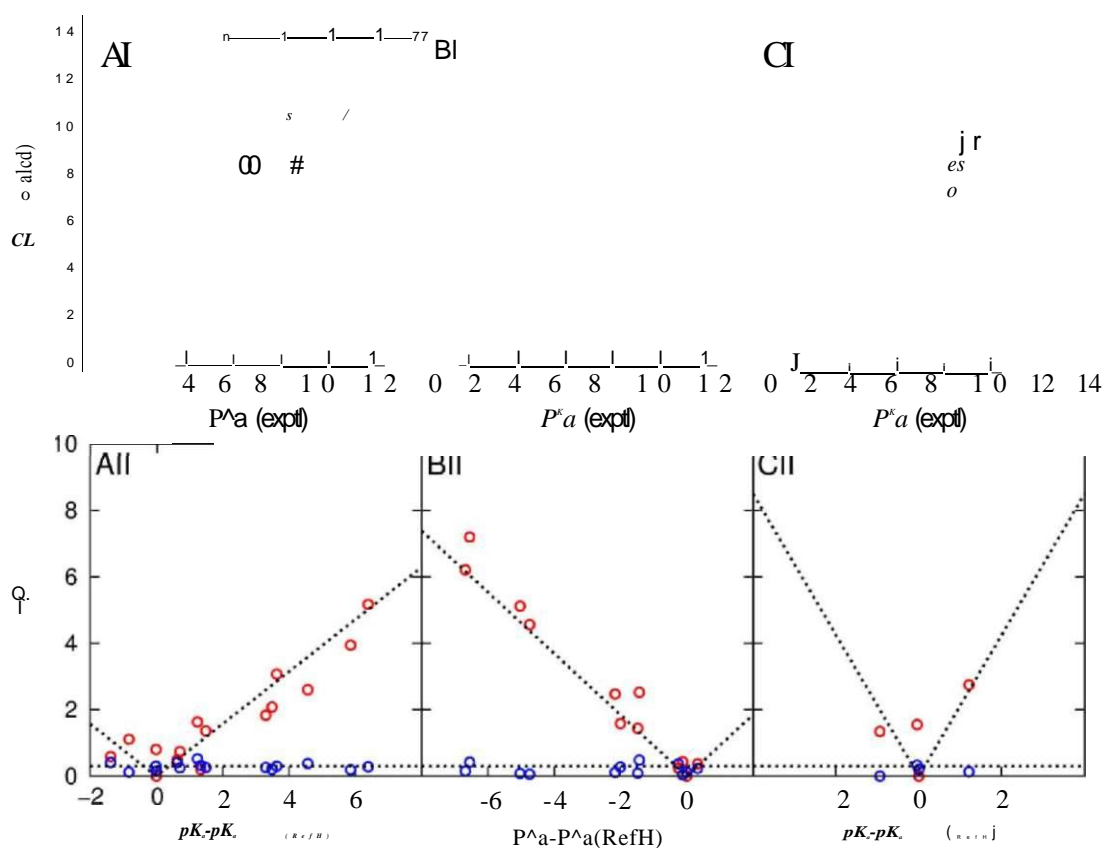


Figure 4. Plot of the predicted  $pK_a$  values with the isodesmic reaction for the pyridine (AI), phenol (BI) and benzylamine (CI) sets against their experimental values. The red and blue circles respectively correspond to the predictions made with the isodesmic reaction before and after the corrections to the calculated energies obtained in the linear regression with the experimental values. The ideal behavior is shown as a reference by the dashed line. Plot of the absolute errors of the predicted  $pK_a$  values against the  $pK_a$  difference with respect to the reference species for pyridines (AII), phenols (BII) and benzylamines (CII). The estimated errors for the uncorrected and corrected predictions are shown as dotted lines.

Finally, the isodesmic reaction approach can be applied to reactions other than proton exchange for the calculation of different thermodynamic constants. Since a possible mechanism of inhibition of advanced glycation end-products by pyridoxamine involves chelation of transition Cu(II) amongst other metals in the organism, it became interesting to study the stability of pyridoxamine-Cu(II) complexes as a function of pH. However, the procedure that will be explained is of general application to complexes formed by any combination of ligand and metal ion. Analogously to the proton exchange schema, the adaptation of the isodesmic reaction to the calculation of stability constants of complexes (logP values) involves a ligand-exchange reaction with respect a reference metal complex species (Results 4.1.4).

A set of Cu(II) complexes formed with ligands of diverse chemical structures was considered to evaluate the performance of the isodesmic reaction. The chosen complexes exhibit ligand-metal bonds through different oxygen, nitrogen and sulphur functionalities, act as monodentate or bidentate ligands, form different chelate ring sizes and exhibit different global charges. Contrarily to the  $pK_a$  predictions, a linear regression of the calculated logP values versus the experimental values results in slope values close to the unit, y-interception close to zero and determination coefficients of 0.96 without statistical corrections. Taking into account that the effects of the continuum solvent are calculated on the solvation spheres of the heavy atoms, it is reasonable that the errors associated to changes of coordination in the metal ion show lower errors than protonation/deprotonation reactions since the exposed area of the metal is much less exposed to the solvent by the presence of the ligands. Moreover, due to the coordination saturation of the metal center the solvent mostly interacts with the ligands, which makes the continuum solvent model a better representation of the real system than in the case of acids. The results indicate that the logP predictions show errors lower than 1.5 units with respect to the experimental values. These errors are larger than in the  $pK_a$  predictions but it should be taken into account that while the experimental uncertainty in the  $pK_a$  values is of 0.01-0.1 units, that of the experimental logP can be easily of 1-1.5 units or higher. Therefore, the errors of the predictions with the isodesmic reaction statistically fall within the experimental uncertainty.

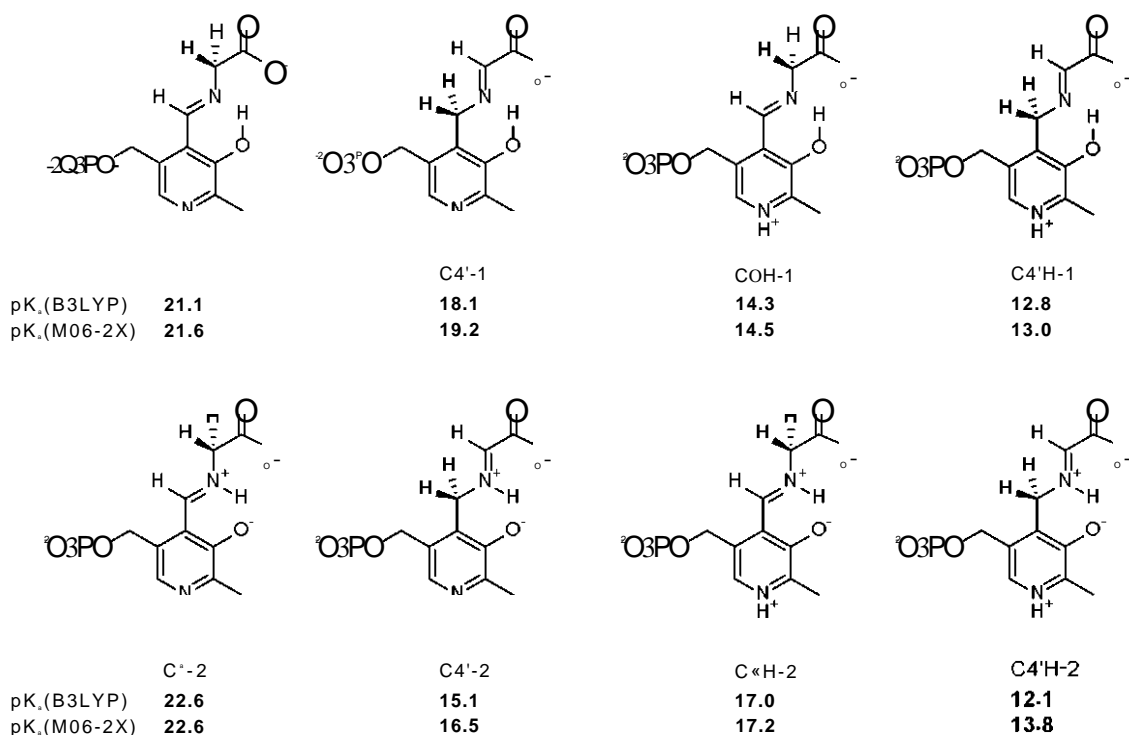
## 5.2. PLP-catalyzed reactions

Protonation and deprotonation events of the Ca carbon of the amino acid or of the C4' carbon of the cofactor, are ubiquitous in the PLP-catalyzed reactions of amino acids. However, only the pK<sub>a</sub> values of the Ca carbon of the major tautomers of the Schiff bases formed between deoxypyridoxal (DPL) and glycine have been experimentally determined. The pK<sub>a</sub> values of the hydrogens bound to carbons are difficult to measure experimentally because their low acidity implies that usually the concentration of dissociated species in aqueous solution is negligible. Other complications may appear. For instance, in the case of PLP and PMP Schiff bases the imine group is unstable both at very basic and acid pH conditions. In addition, PLP and PMP Schiff bases exhibit different tautomeric forms, which may be in low concentrations, in a wide pH range. Finally, the phenol group in the PMP Schiff bases causes the formation of a very stable internal carbinolamine compound that prevents the experimental titration of the C4' carbon in solution. All these limitations make the previous computational strategies an appropriate alternative for the study of the carbon acidities of PLP and PMP Schiff bases. The isodesmic reaction calibrated in a wide pK<sub>a</sub> range will be used since the available pK<sub>a</sub> values of Ca of DPL-Glycine Schiff bases show significant variation in aqueous solution depending on the pH.

PLP and PMP Schiff bases (Scheme 23) are in fact tautomers related by common carbanionic intermediates. It is worth mentioning this tautomeric relationship because the specificity of the reactions catalyzed by PLP-dependent enzymes depends on the protonation of the carbanionic intermediate either at Ca or at C4'. Our results show that PLP Schiff bases, whose Ca is quaternary, are more stable than the PMP tautomers because the conjugation of the imine and hydroxypyridine n-systems is not possible in the latter species due to the quaternary C4' carbon (Results 4.2.2. and 4.2.3.). Considering that the carbanionic intermediates are the conjugate bases of PLP and PMP Schiff bases, a lower stability of the PMP tautomers involves that the C4' carbon is always more acidic than Ca for a particular protonation state (Scheme 23).

The protonation state of the heteroatoms that are integrated in the n system of the imine group or the hydroxypyridine ring strongly modulate the acidity of both Ca and C4' carbons. The most influent heteroatom is the pyridine nitrogen as it stabilizes the carbanionic intermediate by delocalization of the negative charge, which can be formed at Ca or C4', on the n-system of the hydroxypyridine ring. However, the pK<sub>a</sub> reduction of Ca and C4' is significantly different after protonation of the pyridine nitrogen (Results 4.2.2.). For the enolimine tautomers, in which the phenol and imine groups are not ionized (i.e. species **Cx-1**, Scheme 23), protonation of the pyridine nitrogen reduces the pK<sub>a</sub> of Ca by ~7 units (**Ca-1** vs **Ca-1H**) whereas the pK<sub>a</sub> of C4' is reduced by 5-6 units, as observed when comparing **C4'-1** and **C4'-1H**. Protonation of the pyridine nitrogen in the ketoenamine tautomers, which exhibit an ionized phenol and imine groups (i.e. species **Cx-2**, Scheme 23), has lower effects on the pK<sub>a</sub> of both Ca and C4' than those observed for the enolimine tautomers (Results 4.2.2.). As shown in Scheme 23, the pK<sub>a</sub> difference between **Ca-2** and **Ca-2H** is ~5 units, while that of

**C4'-2** and **C4'-2H** is ~3 units. This is due to a destabilization of the carbanionic intermediates of **Cx-2** species as a result of the n-backdonation from the phenoxide anion to the hydroxypyridine-imine conjugate system (Table 1).



Scheme 23. Calculated pK<sub>a</sub> values of the Ca and C4' carbons of the PLP and PMP Schiff bases. The acidic protons are depicted in bold. The labels "1" and "2" stand for the enolimine and ketoenamine tautomers respectively. The label "H" stands for the pyridine nitrogen-protonated Schiff bases.

Regarding the imine-phenol tautomerism, proton transfer from the phenol oxygen to the imine nitrogen decreases the acidity of Ca carbon by 1.5-1.0 pK<sub>a</sub> units as deduced when comparing **Ca-1** and **Ca-2** (Scheme 23). Oppositely, the acidity of the C4' carbon increases by ~3 pK<sub>a</sub> units (**C4'-1** vs **C4'-2**). This difference is indicative of the negative charge delocalization from the phenoxide anion to the cationic imine nitrogen via n-system in **Ca-2** and **Ca-2H** species, which is absent in **C4'-2** and **C4'-2H** species due to the presence of the C4'H<sub>2</sub> methylene group (Table 1). In the latter case, the acidity increase of C4' is caused by destabilization of the C4'-hydrogens due to the adjacent positively charged iminium nitrogen.

The effects of the intramolecular hydrogen bond tautomerism are enhanced when the pyridine nitrogen is protonated since the n-backdonation of the phenoxide group increases. Indeed, the pK<sub>a</sub> of the Ca carbon increases approximately by 3 units from **Ca-1H** to **Ca-2H** instead of the 1.5-1.0 increase observed between **Ca-1** and **Ca-2** (Scheme 23). However, such increase is not observed for the pK<sub>a</sub> of C4' when comparing **C4'-1H** and **C4'-2H** since the destabilization of the C4'-hydrogens by the adjacent iminium cation counteracts the destabilization of the carbanionic intermediate due to the n-backdonation of the phenoxide group.

Table 1. Charge transferred from the imine moiety to the pyridine moiety in the PLP, PMP Schiffbases and their carbanionic intermediates.

Species	Delocalized charge		
Ca-1	0.06		
Ca-1H	0.12		
Ca-2	-0.26		
Ca-2H	-0.19		
C4'-1	0.09		
C4'-1H	0.13		
C4'-2	-0.06		
C4'-2H	-0.03		
anion-1	0.36		
anion-	0.65		
anion-2	0.03		
anion-	0.24		

First, the NBO atomic partial charges of the all atoms were summed up into two groups, namely imine moiety (blue atoms) and pyridine moiety (red atoms). The charge transferred from the imine moiety to the pyridine moiety was calculated as the difference between the summation of partial charges of the imine moiety and the global charge of a hypothetical state with no delocalization. Positive/negative values indicate charge delocalization towards/from the pyridine ring, respectively. The labels anion-x stand for the carbanionic intermediates of the "x" species.

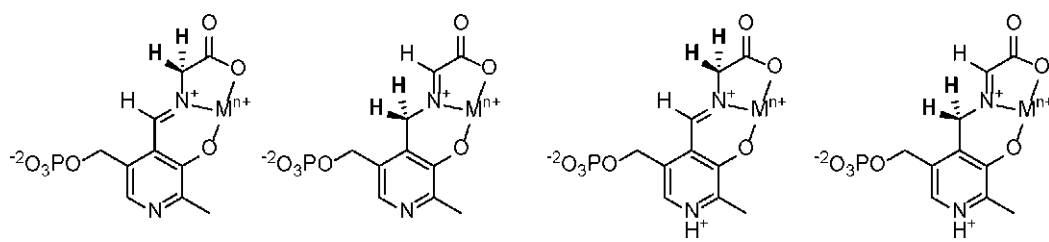
To summarize, protonation of the pyridine nitrogen causes the largest increase in the acidity of Ca and C4'. However, protonation of the imine nitrogen and deprotonation of the phenol oxygen reduce the acidity of Ca and increase that of C4', and its extent is modulated by the protonation state of the pyridine nitrogen. That is, the absolute pK<sub>a</sub> value of Ca and C4' is mainly set by the protonation state of the pyridine nitrogen and the relative pK<sub>a</sub> value between Ca and C4' is set by the protonation state of the imine nitrogen (Results 4.2.2.).

It is well known from experiments that PLP and PMP Schiffbases form stable chelates with metal ions and catalyze a number of reactions on amino acids. However, there is no evidence of PLP-dependent enzymes which also use metals as cofactors. The study of the carbon acidities of Ca and C4' in the metal chelates attempts to shed some light on this apparent contradiction.

As shown in Scheme 24, the carbon acidities of Ca and C4' in the **Ca-M** and **C4'-M** complexes of divalent metals, with the exception of Cu<sup>2+</sup> are very similar to those calculated for the uncomplexed Schiffbases (Scheme 23). In addition, protonation of the pyridine nitrogen in the metal complexes reduces the pK<sub>a</sub> of Ca and C4' by ~7 units (Scheme 24), similarly than for the uncomplexed Schiff bases (Scheme 23).

Amongst the studied divalent metals, only chelation of Cu<sup>2+</sup> increases the acidity of Ca and especially the acidity of C4' with respect to the uncomplexed Schiffbases (Scheme 24).

Chelation of trivalent metals  $\text{Al}^{3+}$  and  $\text{Fe}^{3+}$  has a much larger impact than chelation of divalent metals on the  $\text{pK}_a$  values of Ca and C4' atoms. The Ca carbon in the  $\text{Al}^{3+}$  and  $\text{Fe}^{3+}$  complexes is 4 to 7 units more acidic than in the uncomplexed Schiff bases and divalent metal complexes (Scheme 24). The C4' carbon also exhibits a  $\text{pK}_a$  value that is 5 to 8 units lower than in the uncomplexed Schiff bases (Scheme 24).



$\text{M}^{n+}$	Ca-M	C4'-M	CaH-M	C4'H-M
$\text{Mg}^{2+}$	22.3	17.9	15.8	12.9
$\text{Zn}^{2+}$	21.9	17.4	14.8	11.6
$\text{Ni}^{2+}$	21.7	17.1	15.5	11.0
$\text{Cu}^{2+}$	20.3	12.5	13.1	6.9
$\text{Al}^{3+}$	16.0	9.3	9.4	4.8
$\text{Fe}^{3+}$	16.0	10.2	10.6	4.9

Scheme 24. Calculated  $\text{pK}_a$  values of the Ca and C4' carbons of the PLP and PMP Schiff base metal complexes. The acidic protons are depicted in bold. The M label stands for each specific metal ion in the complexes. The "H" stands for the pyridine nitrogen-protonated Schiff bases.

Based on the calculated carbon acidities of Ca and C4', there is not an obvious reason that explains the absence of metal ions in the active site of PLP-dependent enzymes. However, it is important to note that none of the studied metal ions reduces the  $\text{pK}_a$  of Ca or C4' as efficiently as the protonation of the imine moiety groups. Therefore, the apparent activation of the complexed Schiff bases in solution should be mostly attributed to the stabilization of the Schiff bases, which leads to higher reaction rates as a consequence of a concentration increase. A potential advantage of metal ions is that the complexed Schiff bases are highly stable in a large pH range, which enables unusual protonation states of the Schiff bases that, indirectly, may modify the carbon acidity. However, the protonation state of the Schiff base heteroatoms in the active site of PLP-dependent enzymes is easily modulated by close interactions with polar or charged residues. Taking everything into account, the  $\text{pK}_a$  values of Ca and C4' would be adequate for the catalysis of different reactions in PLP-dependent enzymes, but an excessive stabilization of the Schiff bases could introduce a kinetic disadvantage for the transamination steps (Results 4.2.3.).



The  $^1\text{H}$  NMR experiments studies carried out on the H/D exchange of 4-(aminomethyl)pyridine Schiff bases revealed that the carbanions formed in aqueous solution are moderately stable and the acid-catalyzed reprotonation by dihydrogen phosphate exhibits a non-negligible activation barrier (Results 4.3.2.). It has been reported that the PLP Schiff base carbanions undergo aldol condensation reactions in solution in the presence of carbonyl species. These results promoted a more detailed study on the kinetic competence between protonation/deprotonation reactions at Ca and aldol condensation reactions with sugars as a possible route of carbonyl scavenging and inhibition of glycation reactions (Results 4.3.1.).

The calculated activation free energies of acid-catalyzed reprotonation of Ca are higher than that of aldol condensation between the Ca carbanion and glyceraldehyde in all cases with the exception of the acetic acid-catalyzed reprotonation. However, this reaction is negligible at physiological pH due to the absence of protonated carboxylic acids. Additionally, the activation barrier for the reprotonation reaction involving water solvent is approximately  $\sim 7$  kcal/mol, which supports that the formed carbanions have long enough half-life times to act as carbonyl scavengers. Furthermore, the reduced carbonyl group condensed with the amino acid is a thermodynamically stable product that can be hydrolyzed, regenerating the PLP catalyst (Results 4.3.1.).

According to the free energy profiles obtained for the different proposed mechanisms, the evolution of the towards the final hydrolysis products through a stable carbinolamine is also kinetically favored. The 5'-phosphate group favors the addition of the carbanion to the carbonyl group in the aldol condensation step and in the release of the end-products in the final step of the hydrolysis. However, this group hinders the formation of the carbinolamine intermediate due to stabilization of its precursor state. Taking into account the high mobility of the 5'-phosphate group in solution at room temperature, it is reasonable that only participates in the most favored steps (Results 4.3.1.).

At last, the decarboxylation reaction of ornithine is the second studied type of Ca-C reaction catalyzed by PLP. According to the metadynamics simulations and ONIOM calculations, Ornithine decarboxylase accelerates the elimination rate of the carboxylate group by a  $10^{19}$ -fold increase (Results 4.3.2.). However, the simulations in solution show that the PLP cofactor is responsible for  $10^{16}$  of the total catalysis and, therefore, the exclusive contribution of the enzyme only constitutes the remaining 10 - fold acceleration (Results 4.3.2.). The enzymatic environment reduces the direct interactions with the carboxylate group almost to the maximum and reduces the activation barrier by destabilization of the reactants with respect to the transition state. In addition, the enzyme structure creates an electrostatic gradient which destabilizes the reactants by placing more anionic than cationic residues closer to the carboxylate group (Results 4.3.2.). It should be noted that the enzyme promotes the protonation of the PLP pyridine nitrogen in the active site, which accounts for  $10^6$  of the catalysis by the cofactor, but it is deprotonated at physiological pH in solution (Results 4.3.2.).

Taking into account the data obtained for the acidities of Ca and C4', the acid-catalyzed protonation/deprotonation activation barriers on C a, and the activation

barriers for the aldol condensation and decarboxylation reactions at Ca, interesting conclusions on the reaction specificity of PLP-dependent enzymes can be drawn.

In the active site of Alanine racemase (AlaR), the pyridine nitrogen of the PLP-Alanine Schiff base interacts with the guanidinium group of Arg219 (Figure 5). Considering that the pyridine is much more acidic than the guanidinium group, the pyridine nitrogen will be deprotonated in the active site. An additional guanidinium group of a second arginine, Arg136, establishes short hydrogen bonding interactions with the phenol and carboxyl oxygens (Figure 5), which stabilize these groups as anions. As shown in Figure 5, the  $\epsilon$ -amino group of Lys39 is almost equidistant to Ca and C4'. Therefore, once Ca is deprotonated by Tyr265' and the carbanionic intermediate is formed, the position of the Lys39  $\epsilon$ -amino group does not particularly favor the reprotonation at Ca in front of at C4', which potentially increases the probability of transamination side-reactions by reprotonation at C4'.

According to the calculated carbon acidities (Scheme 23), the protonation state of the PLP-Alanine Schiff base in the active site is the least favorable for the activation of the Ca-H hydrogen. However, the  $pK_a$  difference between Ca and C4' is maximized since almost no negative charge is delocalized from Ca to the ring (Table 1). Therefore, our results indicate that high activation of Ca is deterred in favor of high reaction specificity towards racemization.

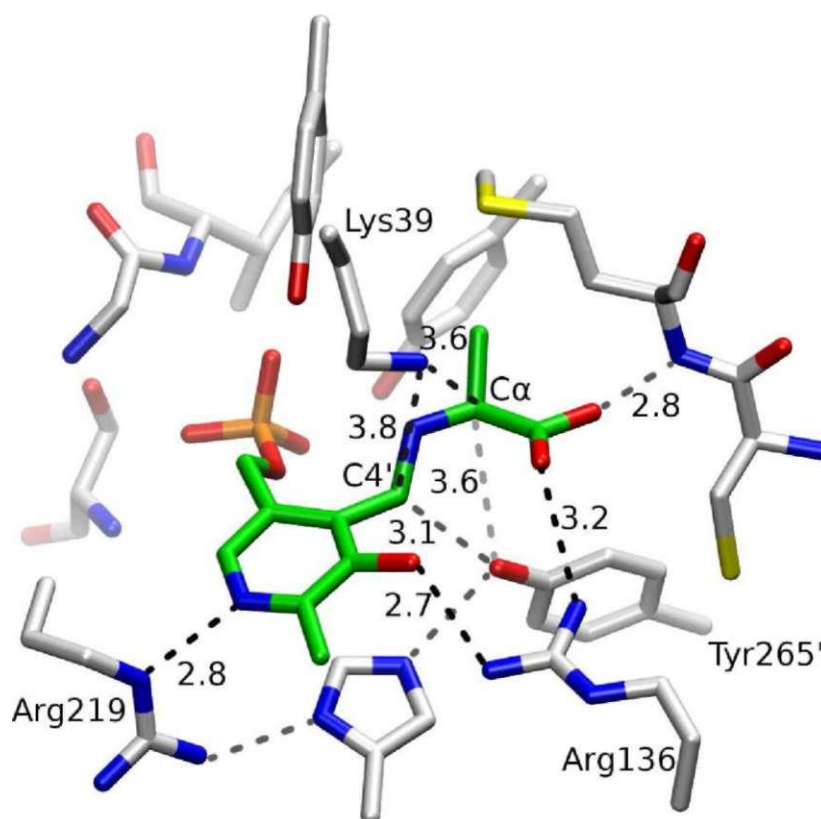


Figure 5. Active site interactions of the PLP-Alanine Schiff base in the active site of Alanine racemase (PDB code 1L6F). All distances are measured in Angstrom.

In the case of Aspartate aminotransferase (AAT), the main reaction proceeds via deprotonation of Ca and reprotonation of the carbanionic intermediate at C4'. An inspection of the crystallographic active site reveals that, oppositely to Alanine racemase, the pyridine nitrogen of the PLP-Aspartate Schiff base interacts with an acidic residue (Asp222, Figure 6). This interaction points out that the pyridine nitrogen is protonated in this active site. The phenol oxygen forms hydrogen bonding interactions with neutral residues (Figure 6), which do not favor the proton transfer from the phenol oxygen to the imine nitrogen by electrostatic stabilization of the resultant phenoxide anion. However, it should be noted that the protonated pyridine nitrogen contributes to stabilize the deprotonated phenoxide group.

The present pKa calculations indicate that the AAT active site promotes the protonation of the pyridine nitrogen to activate the most the Ca carbon for deprotonation. In addition, the lack of cationic residues and extensive hydrogen bonding interactions at the phenol oxygen suggest that it remains protonated or that the imine nitrogen transfers the proton to the phenoxide anion after the formation of the carbanionic intermediate. Accordingly, not only does the AAT active site promote the activation of Ca, but also reduces the pKa difference between Ca and C4' favoring the reprotonation at this last position as a result.

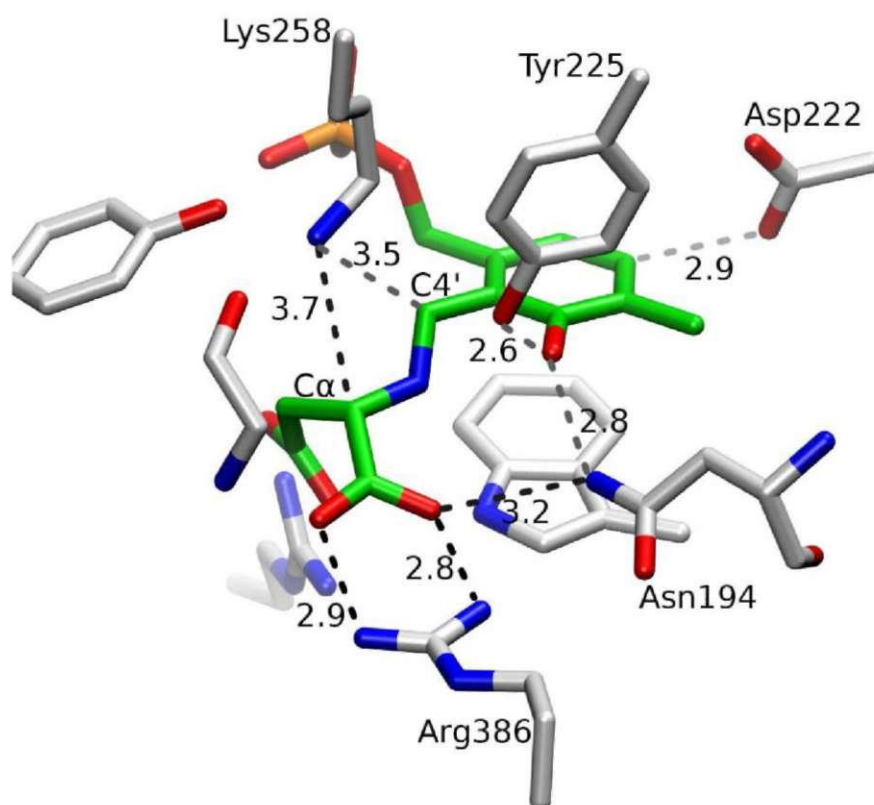
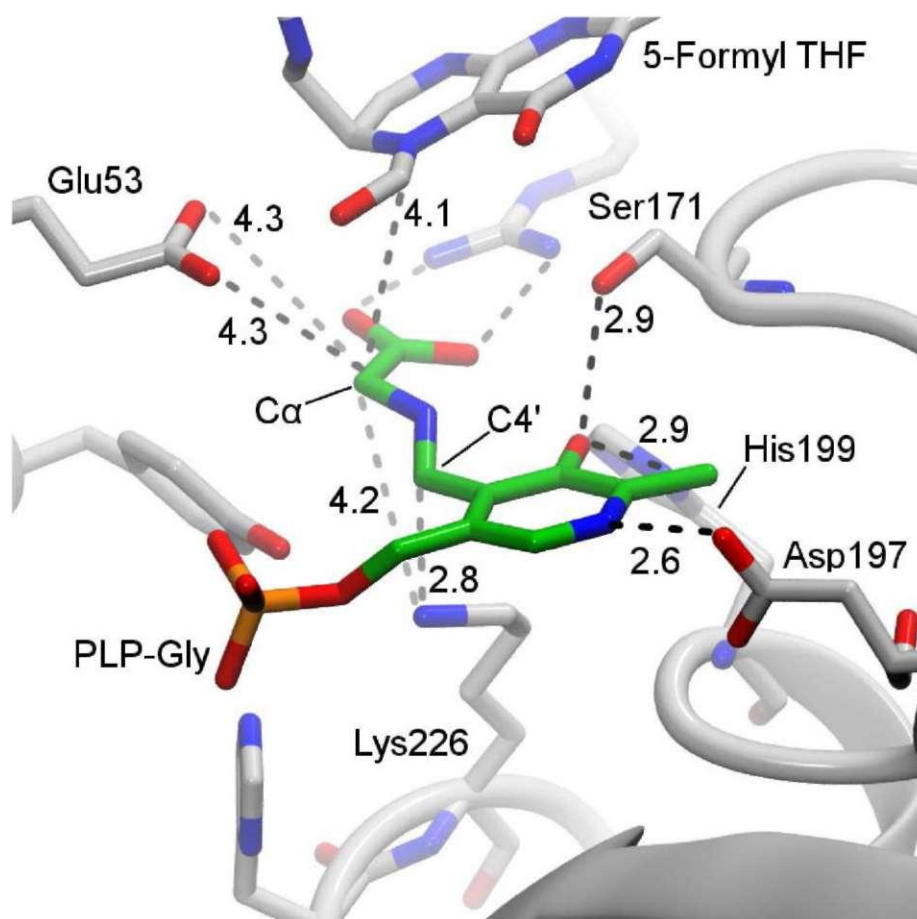


Figure 6. Active site interactions of the deaza-PLP-Aspartate Schiff base in the active site of Aspartate aminotransferase (PDB code 3QPG). All distances are measured in Angstrom.



Similarly to the active site of Aspartate aminotransferase, the residues interacting with the phenol oxygen O3' are neutral (Ser171 and His199) (Figure 7), which are not as effective as cationic residues to stabilize the phenoxide anion. Besides, the Schiff base pyridine nitrogen is presumably protonated since its closest interacting residue is the sidechain carboxyl group of Asp197 (Figure 7). Finally, the e-amino group of Lys226 is closer to the C4' carbon than to the Ca one (Figure 7).

Analogously to the reaction in Aspartate aminotransferase, these interactions with the Schiff base suggest a more favorable reprotonation of the carbanionic intermediate at C4'. Therefore, Lys226 is unlikely to be the acid catalyst in the reprotonation reaction. As seen in Figure 7, there is a carboxylate group of a glutamic acid residue (Glu53) close to the Ca carbon. According to the calculated activation barriers of reprotonation and the activation barriers of aldol condensation, only the reprotonation of Ca catalyzed by carboxylic acids are faster than the Ca-C bond formation in the aldol condensation reaction (Results 4.3.1.). Thus the carboxyl sidechain of Glu53 is probably the actual catalytic acid acting in the reprotonation step of the carbanionic intermediate, ensuring the protonation at Ca.

In the active site of Ornithine decarboxylase, the interactions displayed by the Schiff base are very similar to those observed for the Aspartate aminotransferase and Serine hydroxymethyltransferase active sites.

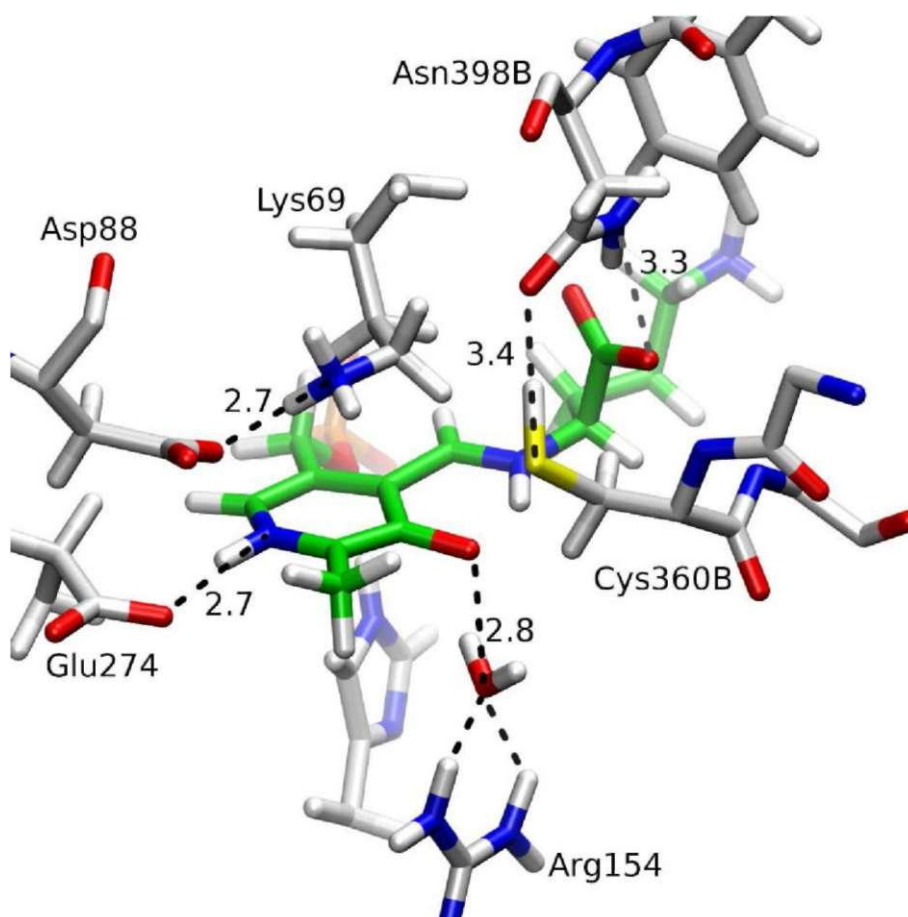


Figure 8. Interactions of the PLP-Ornithine Schiff base in the active site of Ornithine decarboxylase (Molecular dynamics snapshot of one active site of ODC reconstructed from PDB code 2O00). All distances are measured in Angstrom.

As shown in Figure 8, the pyridine nitrogen is protonated by interactions with the sidechain of a glutamic residue (Glu274), and the phenol oxygen only interacts with a water molecule. In addition, the average distances calculated in the simulations show that the s-amino group of Lys69 is significantly closer to C4' than to Ca (Results

4.3.2.). Considering that such a protonation state of the PLP Schiff base will promote the charge delocalization of the carbanionic intermediate to the pyridine ring, reprotonation of such intermediate via acid catalysis by Lys69 would clearly be favored at C4' instead of Ca, which in turn would lead to transamination-side reactions. However, only protonation at Ca is observed for the native enzyme.

According to the performed metadynamics simulations, desolvation of the Schiff base is restricted in the active site to maintain a high energy carbanionic intermediate, which also hinders full proton transfer from the imine nitrogen to the phenol oxygen (Results 4.3.2.). This mechanism reduces transamination side-reactions by avoiding a maximum delocalization of the negative charge from Ca to the pyridine ring. In addition, elimination of CO<sub>2</sub> after decarboxylation liberates the amide group of Asn398B, which in turn releases the thiol group of Cys360B allowing its rotation towards the active site. As has been suggested from X-ray experiments, when the thiol group of Cys360B is rotated, it is at a correct distance from Ca for the reprotonation step (Figure 9). This hypothesis is supported by the activation energies of reprotonation at Ca calculated in 4.3.1, which indicate that thiol and amine groups are equally effective acid catalysts. Thus, Lys69 is not required for the reprotonation at Ca, which reduces the probability of accidental reprotonation events at C4' and the consequential transamination-side reactions.

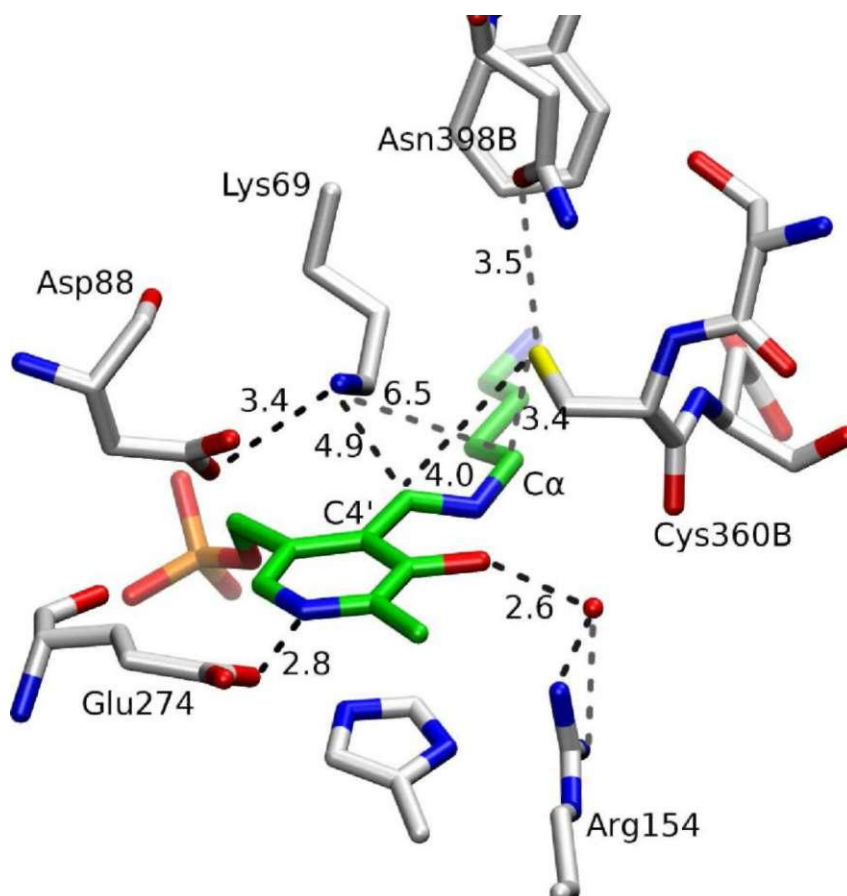


Figure 9. Interactions of the PLP-Putrescine Schiff base in the active site of Ornithine decarboxylase (PDB code 1F3T). All distances are measured in Angstrom.







## **6. Conclusions**



- 1- The precision of the  $pK_a$  predictions when using thermodynamic cycles is strongly dependent on the cancellation of errors of the calculated solvation free energies. The introduction of explicit water molecules improves the  $pK_a$  predictions by reducing the errors of the electrostatic interaction between the dielectric continuum with the more charged atoms of the solute.
- 2- The isodesmic reaction scheme maximizes the cancellation of errors by calculating proton exchange free energies in solution and avoiding gas-phase calculations. A set of molecules and their  $pK_a$  values can be used to correct the calculated proton exchange free energies and reduce the errors of the predictions in a large  $pK_a$  range.
- 3- Schiff base formation with PLP reduces the  $pK_a$  of the Ca of amino acids to values that are typical of moderately strong acids in solution. The largest contribution to the  $pK_a$  reduction is the protonation of the pyridine nitrogen of the PLP cofactor. However, the relative acidity between the Ca and the C4' carbons is mostly determined by the protonation state of the imine nitrogen and the phenol oxygen
- 4- The protonation state of the pyridine, imine and phenol groups is crucial to determine the evolution of the carbanionic intermediates in all PLP-catalyzed reactions. X-ray structures support that each PLP-dependent enzyme modulates the protonation state of PLP Schiff bases to achieve high reaction specificity for the desired reaction.
- 5- The extraordinary decarboxylation rate catalyzed by Ornithine decarboxylase is mostly due to the PLP cofactor. The catalytic contribution exclusive to Ornithine decarboxylase is due to ground-state destabilization in the active site by desolvation of the carboxylate group and by an unfavourable electrostatic potential.
- 6- Ornithine decarboxylase reduces the activation energy of the decarboxylation step so that it is not rate-limiting and simultaneously destabilizes the formed carbanionic intermediate to ensure its correct protonation at the Ca carbon.



## **7. Bibliography**



- (Adrover2005) Adrover, M.; Vilanova, B.; Muñoz, F.; Donoso J. *Chem. Biodiv.* **2005**, 2, 964-974.
- (Adrover2007) Adrover, M.; Vilanova, B.; Muñoz, F.; Donoso J. *Int. J. Chem. Kin.* **2007**, 3, 154-167.
- (Adrover2008) Adrover, M.; Vilanova, B.; Frau, J.; Muñoz, F.; Donoso, J. *Bioorg. Med. Chem.* **2008**, 16, 5557-5569.
- (Adrover2009) Adrover, M.; Vilanova, B.; Frau, J.; Muñoz, F.; Donoso J. *Amino Acids* **2009**, 36, 437-448.
- (Bach1997) Bach, R. D.; Canepa, C. *J. Am. Chem. Soc.* **1997**, 119, 11725-11733.
- (Bach1999) Bach, R. D.; Canepa, C.; Glukovtsev, M. N. *J. Am. Chem. Soc.* **1999**, 121, 6542-6555.
- (Bader1990) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*, Claredon Press, Oxford **1990**.
- (Bailey1998) Bailey, A. J.; Paul, R. G.; Knott, L. *Mech. Ageing Dev.* **1998**, 106, 1-56.
- (Barducci2011) Barducci, A.; Bonomi, M.; Parrinello, M. *John Wiley & Sons, Ltd. WIREs Comput. Mol. Sci.* **2011**, 1, 826-843.
- (Barone1997) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, 107, 3210-3221.
- (Barone1998) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, 102, 1995-2001.
- (Baynes1999) Baynes, J. W.; Thorpe, S. R. *Diabetes* **1999**, 48, 1-9.
- (Biegler-König2001) AIM2000, Biegler-König, F.; Schönbohm, J. Innovative software, in F. Biegler-König, Schömböhm, J.; Bayles (Eds.) *J. Comp. Chem.* **2001**, 22, 545-559.
- (Breneman1990) Breneman, C. M.; Wiberg, K. B. *J. Comp. Chem. A* **1990**, 11, 361-373.
- (Brooks1997) Brooks, H.; Phillips, M. A. *Biochemistry* **1997**, 36, 15147-15155.
- (Brownlee1992) Brownlee, M. *Diabetis Care* **1992**, 15, 1835-1843.
- (Brzovic1990) Brzovic, P.; Holbrook, E. L.; Greene, R. C.; Dunn, M. F. *Biochemistry* **1990**, 29, 442-451.
- (Caldes2011) Caldés, C.; Vilanova, B.; Adrover, M.; Muñoz, F.; Donoso, J. *Bioorg. Med. Chem.* **2011**, 19, 4536-4543.

- (Cerqueira2011) Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. *J. Chem. Theory Comput.* **2011**, *7*, 1356-1368.
- (Chan-Huot2010) Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H. H. *Biochemistry* **2010**, *49*, 10818-10830.
- (Christen1985) Christen, P.; Metzler, D. E. In *Transaminases*; John Wiley & Sons: New York, **1985**.
- (Cossi2003) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669-681.
- (Cramer2004) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*, Second ed., John Wiley and Sons, Chichester, **2004** (Chapters 10, 11).
- (Crueiras2008) Crueiras, J.; Rios, A.; Riveiros, E.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2008**, *130*, 2041-2050.
- (Crueiras2009) Crueiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 15815-15824.
- (Dunathan1966) Dunathan, H.C. *Proc. Natl. Acad. Sci. USA* **1966**, *55*, 712-716.
- (Eliot2004) Eliot, A. C.; Kirsch, J. F.; *Annu. Rev. Biochem.* **2004**, *73*, 383-415.
- (Evangelopoulos1984) Evangelopoulos, A. E. In *Chemical and Biological Aspects of Vitamin B6 Catalysis*; Liss, R. A., Ed.; New York, **1984**.
- (Fogle2011) Fogle, E. J.; Toney, M. D. *Biochimica et Biophysica Acta* **2011**, *1814*, 1113-1119.
- (Glendening) NBO Version 3.1, Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F.
- (Griswold2012) Griswold, W. R.; Nieto Castro, J.; Fisher, A. J.; Toney, M. D. *J. Am. Chem. Soc.* **2012**, *134*, 8436-8438.
- (Grillo2008) Grillo, M. A.; Colombatto, S. *Amino Acids* **2008**, *35*, 29-36.
- (G09) Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.;



Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

(Hoyer2002) Hoyer, S. J. *J. Neural. Transm.* **2002**, *109*, 341-360.

(Jackson2000) Jackson, L. K.; Brooks, H. B.; Osterman, A. L.; Goldsmith, E. J.; Philips, M. A. *Biochemistry* **2000**, *39*, 11247-11257.

(Jackson2003) Jackson, L. K.; Brooks, H. B.; Myers, D. P.; Philips, M. A. *Biochemistry* **2003**, *42*, 2933-2940.

(Jansonius1998) Jansonius, J. N. *Curr. Opin. Struct. Biol.* **1998**, *8*, 759-769.

(Jomova2010) Jomova, K.; Vondrakova, D.; Lawson, M.; Valko, M. *Mol. Cell. Biochem.* **2010**, *345*, 91-104.

(Kallies1997) Kallies, B.; Mitzner, R. *J. Phys. Chem. B* **1997**, *101*, 2959-2967.

(Kepp2012) Kepp, K. P. *Chem. Rev.* **2012**, *112*, 5193-5239.

(Kiruba2003) Kiruba, G. S. M.; Wong, M. W. *J. Org. Chem.* **2003**, *68*, 2874-2881.

(Klamt1993) Klamt, A.; Schüürmann G. *J. Chem. Soc. Perkin Trans. 2* **1993**, *5*, 799-805.

(Kume1995) Kume, S.; Takeya, M.; Mori, T.; Araki, N.; Suzuki, H.; Horiuchi, S.; Kodama, T.; Miyauchi, Y.; Takahashi, K. *Am. J. Pathol.* **1995**, *147*, 654-667.

(Laio2002) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562-12566.

(LeMagueres2005) LeMagueres, P.; Im, J.; Ebalunode, J.; Strych, U.; Benedik, M.J.; Briggs, J. M.; Kohn, H.; Krause, K. L., *Biochemistry* **2005**, *44*, 1471-1481.

(Leussing1986) Leussing, D. L. in *Vitamin B6 Pyridoxal phosphate. Chemical, Biochemical, and Medical Aspects*. Chap. 4, Dolphin, D.; Poulson, R.; Avramovic, O. Eds.; John Wiley & Sons: New York, **1986**.

(Li2008) Li, Y.; Cohenford, M. A.; Dutta, U.; Dain, J. A. *Anal. Bioanal. Chem.* **2008**, *390*, 679-688.

(Lin2010) Lin, Y.; Gao, J. *Biochemistry* **2010**, *49*, 89-94.

(Lin2011) Lin, Y.; Gao, J. *J. Am. Chem. Soc.* **2011**, *133*, 4398-4403.

- (Liptak2001) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, 55, 7314-7319.
- (Liptak2001/2) Liptak, M. D.; Shields, G. C. *Int. J. Quant. Chem.* **2001**, 55, 727-741.
- (Liptak2004) Liptak, M. D.; Gros, K. C.; Seybold, P. G.; Feldgus, G. C.; Shields, G. C. *J. Am. Chem. Soc.* **2004**, 725, 6421-6427.
- (Livanova2002) Livanova, N. B.; Cheborateva, N. A.; Eronina, T. B.; Kurganov, B. I. *Biochemistry* **2002**, 76, 1089-1098.
- (Lopez2002) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, 726, 5010-5018.
- (Lu2003) Lu, H.; Chen, X.; Chang-Guo, Z. *J. Phys. Chem. B* **2003**, 777, 10599-10605.
- (Maillard1912) Maillard, L. C. *C. R. Hebd. Seances Acad. Sci.* **1912**, 754, 66-68.
- (Major2006) Major, D. T.; Nam, K.; Gao, J. *J. Am. Chem. Soc.* **2006**, 725, 8114-8115.
- (Major2006/2) Major, D. T.; Gao, J. *J. Am. Chem. Soc.* **2006**, 725, 16345-16357.
- (Martell1989) Martell, A. E. *Acc. Chem. Res.* **1989**, 22, 115-124.
- (Martell1984) Martell, A. E.; Taylor, P. *Inorg. Chem.* **1984**, 23, 2734-2735.
- (McQuarrie1970) McQuarrie, D. M. (Ed.), *Statistical Mechanics*, Harper and Row, New York, **1970**.
- (Miles2001) Miles, E. W. *Chem. Rec.* **2001**, 7, 140-151.
- (Miranda2010) Miranda, H. V.; Outeiro, T. F. *J. Pathol.* **2010**, 277, 13-25.
- (Miyazawa2012) Miyazawa, T.; Nakagawa, K.; Shimasaki, S.; Nagai, R. *Amino Acids*, **2012**, 42, 1163-1170.
- (Montgomery1994) Montgomery Jr., J. A.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1994**, 707, 5900-5909.
- (Montgomery2000) Montgomery Jr., J. A.; Frisch M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, 772, 6532-6542.
- (Montgomery1999) Montgomery Jr., J. A.; Frisch M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, 770, 2822-2827.
- (Namazian2003) Namazian, M.; Heidary, H. *J. Mol. Struct. THEOCHEM* **2003**, 620, 257-263.
- (Nyden1981) Nyden, M. R.; Petersson, G. A. *J. Chem. Phys.* **1981**, 75, 1843-1862.

- (Ochterski1996) Ochterski, J. W.; Petersson, G. A.; Montgomery Jr., J. A. *J. Chem. Phys.* **1996**, 104, 2598-2619.
- (Oliveira2011) Oliveira, E. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. *J. Am. Chem. Soc.*, **2011**, 133, 15496-15505.
- (Ortega-Castro2009) Ortega-Castro, J.; Adrover, M.; Frau, J.; Donoso, J.; Muñoz, F. *Chem. Phys. Lett.* **2009**, 475, 277-284.
- (Ortega-Castro2010) Ortega-Castro, J.; Adrover, M.; Frau, J.; Salva, A.; Donoso, J.; Muñoz, F. *J. Phys. Chem. A* **2010**, 114, 4634-4640.
- (Ortega-Castro2012) Ortega-Castro, J.; Frau, J.; Casasnovas, R.; Fernández, D.; Donoso, J.; Muñoz, F. *J. Phys. Chem. A* **2012**, 116, 2961-2971.
- (Palascak2004) Palascak, M. W.; Shields, G. C. *J. Phys. Chem A* **2004**, 105, 3692-3694.
- (Percudani2003) Percudani, R.; Peracchi, A. *EMBO Rep.* **2003**, 4, 850-854.
- (Petersson1988) Petersson, G. A.; Bennett, A.; Tensfeld, T. G.; Al-Laham, M. A.; Shirley, W.; Matzaris, J.; *J. Chem. Phys.* **1988**, 89, 2193-2218.
- (Petersson1991) Petersson, G. A.; Al-Laham, M. A. *J. Chem. Phys.* **1991**, 94, 6081-6090.
- (Petersson1983) Petersson, G. A.; Yee, A. K.; Bennett, A. *J. Chem. Phys.* **1983**, 83, 5105-5126
- (Pickard2006) Pickard IV, F. C.; Griffith, D. R.; Ferrara, S. J.; Liptak, M. D.; Kirschner, K. N.; Shields, G. C. *Int. J. Quantum Chem.* **2006**, 106, 3122-3128.
- (Pliego2000) Pliego Jr., J. R.; Riveros, J. M. *Chem. Phys. Lett.* **2000**, 332, 597-602.
- (Pokon2001) Pokon, E. K.; Liptak, M. D.; Feldgus, S.; Shields, G. C. *J. Phys. Chem. A* **2001**, 105, 10483-10487.
- (Rabbani2012) Rabbani, M.; Thornalley, P. J. *Amino Acids*, **2012**, 42, 1087-1096.
- (Ramasamy2005) Ramasamy, R.; Vannucci, S. J.; Shi, D. Y. S.; Herold, K.; Shi, F. Y.; Schmidt, A. M. *Glycobiology* **2005**, 15, 16-28.
- (Takano2005) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, 1, 70-77.
- (Thornalley1984) Thornalley, P.; Wolff, S.; Crabbe, J.; Stern, A. *Biochim. Biophys. Acta* **1984**, 797, 276-287.
- (Toney2001) Toney, M. D. *Biochemistry* **2001**, 40, 1378-1384.

- (Toney2005) Toney, M. D. *Arch. Biochem. Biophys.* **2005**, 433, 279-287.
- (Toney2011) Toney, M. D. *Biochimica et Biophysica Acta* **2011**, 1814, 1407-1418.
- (Toth2001) Toth, A. M.; Liptak, M. D.; Philips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, 114, 4595-4606.
- (Radzicka1996) Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, 118, 6105-6109.
- (Reed1988) Reed, A. E.; Curtis, L. A.; Weinhold, F. *Chem. Rev.* **1988**, 88, 899-926.
- (Richard2002) Richard, J. P.; Williams, G.; O'Donoghue, A. C.; Amyes, T. L. *J. Am. Chem. Soc.* **2002**, 124, 2957-2968. Rios, A.; Richard, J. P.; Amyes, T. L. *J. Am. Chem. Soc.* **2002**, 124, 8251-8259.
- (Richard2009) Richard, J. P.; Amyes, T. L.; Crueiras, J.; Rios, A. *Curr. Op. Chem. Biol.* **2009**, 13, 475-483.
- (Rios1997) Rios, A.; Richard, J. P. *J. Am. Chem. Soc.* **1997**, 119, 8375-8376.
- (Rios2000) Rios, A.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2000**, 122, 9373-9385.
- (Rios2001) Rios, A.; Crueiras, J.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2001**, 123, 7949-7950.
- (Salva2001) Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. *J. Mol. Struct. THEOCHEM* **2001**, 577, 229-238.
- (Salva2002) Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. *Int. J. Quantum Chem.* **2002**, 89, 48-56.
- (Salva2003) Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. *J. Phys. Chem. A* **2003**, 107, 9409-9414.
- (Salva2004) Salva, A.; Donoso, J.; Frau, J.; Muñoz, F. *J. Phys. Chem. A* **2004**, 108, 11709-11714.
- (Schirch2005) Schirch, V.; Szebenyi, M. E. *Curr. Op. Chem. Biol.* **2005**, 9, 482-487.
- (Shaw1997) Shaw, P. J.; Petsko, G. A.; Ringe, D. *Biochemistry* **1997**, 36, 1329-1342.
- (Snider2000) Snider, M. J.; Wolfenden, R. *J. Am. Chem. Soc.* **2000**, 122, 11507-11508.
- (Spies2007) Spies, M. A.; Toney, M. D. *J. Am. Chem. Soc.* **2007**, 129, 10678-10685.
- (Spies2004) Spies, M. A.; Woodward, J. J.; Watnik, M. R.; Toney, M. D. *J. Am. Chem. Soc.* **2004**, 126, 7464-7475.

- (Sun1999) Sun, S.; Toney, M. D. *Biochemistry* **1999**, *38*, 4058-4065.
- (Szebenyi2004) Szebenyi, M. E.; Musayev, F. N.; di Salvo, M. L.; Safo, M. K.; Schirch, V. *Biochemistry* **2004**, *43*, 6865-6876.
- (Tissandier1998) Tissandier, M. D.; Cowe, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, *102*, 7787-7794.
- (Toney2001) Toney, M. D. *Biochemistry* **2001**, *40*, 1378-1384.
- (Toth2007) Toth, K.; Richard, J. P. *J. Am. Chem. Soc.* **2007**, *129*, 3013-3021.
- (Vazquez1989) Vázquez, M. A.; Echevarría, G.; Muñoz, F.; Donoso, J.; García-Blanco, F. *J. Chem. Soc. Perkin Trans. 2* **1989**, *11*, 1617-1622.
- (Vazquez1990) Vázquez, M. A.; Muñoz, F.; Donoso, J.; García-Blanco, F. *Int. J. Chem. Kinet.* **1990**, *22*, 905-914.
- (Vazquez1991) Vázquez, M. A.; Muñoz, F.; Donoso, J.; García-Blanco, F. *J. Chem. Soc. Perkin Trans. 2* **1991**, *2*, 275-281.
- (Vazquez1992) Vázquez, M. A.; Muñoz, F.; Donoso, J.; García-Blanco, F. *Amino Acids* **1992**, *3*, 81-94.
- (Vilanova2004) Vilanova, B.; Adrover, M.; Muñoz, F.; Donoso, J. *Chemistry & Biodiversity* **2004**, *1*, 1073-1089.
- (Voziyan2002) Voziyan, P. A.; Metz, T. O.; Baynes, J. W.; Hudson, B. G. *J. Biol. Chem.* **2002**, *277*, 3397-3403.
- (Voziyan2005) Voziyan, P. A.; Hudson, B. G. *Cell. Mol. Life Sci.* **2005**, *62*, 1671-1681.
- (Weng1983) Weng, S.; Leussing, D. L. *J. Am. Chem. Soc.* **1983**, *105*, 4082-4090.
- (Wolfenden2001) Wolfenden, R.; Snider, M. *J. Acc. Chem. Res.* **2001**, *34*, 938-945.
- (Zabinski2001) Zabinski, R. F.; Toney, M. D. *J. Am. Chem. Soc.* **2001**, *123*, 193-198.

Additionally the following textbooks and manuals have been followed for the writing of the Introduction and Methodology sections:

*Enzymatic Reaction Mechanisms.* Walsh, C. Ed. W. H. Freeman and Company, New York, **1977**

*Transaminases. Biochemistry: A Series of Monographs.* Christen, P.; Metzler, D. E. Eds.; John Wiley & Sons: New York, **1985**

*Vitamin B6 Pyridoxal phosphate. Chemical, Biochemical, and Medical Aspects.* Dolphin, D.; Poulson, R.; Avramovic, O. Eds.; John Wiley & Sons: New York, **1986**

*Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Szabo, A.; Ostlund, N. S. Dover Publications, Mineola, New York, **1996**

*Essentials of Computational Chemistry, Theories and Models 2<sup>nd</sup> Ed.* Cramer, C. J. John Wiley & Sons: New York, **2004**

*Theoretical and Computational Chemistry: Foundations, Methods and Techniques.* Andrés, J.; Bertran, J. Eds.; Universitat Jaume I, **2007**

*Introduction to Computational Chemistry 2<sup>nd</sup> Ed.* Jensen, F. John Wiley & Sons: New York, **2007**







Pyridoxal phosphate (PLP) is a cofactor of more than a hundred enzymes that catalyze amino acid reactions such as racemizations, transaminations, aldol condensations and decarboxylations amongst others. All the PLP-catalyzed reactions entail, at least, one step in which the C $\alpha$  carbon of the amino acid or the C4' carbon of the PLP is deprotonated or protonated. Furthermore, protonation of the formed carbanion intermediates is the common crossroad to all possible reactions and determines the final products. However, the experimental determination of carbon acidities is difficult and involves large uncertainties.

The present work proposes several computational strategies that combine ab initio calculations and solvent models for the accurate prediction of  $pK_a$  values. Such methodology is used to study the thermodynamics of C-H activation in PLP Schiff bases, which is complemented from the kinetic standpoint with the calculation of carbon protonation/deprotonation activation free energies. The PLP-catalyzed aldol condensation of amino acids and sugars is studied in relation with its influence in the inhibition of glycation reactions of biomolecules. Finally, metadynamics simulations are performed to understand the origins of PLP-catalyzed decarboxylation reactions in enzymes and in aqueous- and gas-phases. Altogether, the present results explain how PLP-dependent enzymes control the specificity of the desired reaction by favoring specific protonation states of the PLP Schiff bases.

